Towards Cyber-secure Intelligent Electrical Power Grids
Vulnerability Analysis and Attack Detection

Pan, Kaikai

DOI
10.4233/uuid:4b4f9f96-237e-421b-82f4-97b1393ae507

Publication date
2020

Document Version
Final published version

Citation (APA)
Pan, K. (2020). *Towards Cyber-secure Intelligent Electrical Power Grids: Vulnerability Analysis and Attack Detection*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:4b4f9f96-237e-421b-82f4-97b1393ae507

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

TOWARDS CYBER-SECURE INTELLIGENT ELECTRICAL POWER GRIDS

VULNERABILITY ANALYSIS AND ATTACK DETECTION

TOWARDS CYBER-SECURE INTELLIGENT ELECTRICAL POWER GRIDS

VULNERABILITY ANALYSIS AND ATTACK DETECTION

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Thursday 12 March 2020 at 10:00 o'clock

by

Kaikai PAN

Master of Science in Instrumentation Engineering
Beihang University, Beijing, China
born in Pan'an, Zhejiang Province, China

This dissertation has been approved by the promotors.

promotor: Prof.dr. P. Palensky
copromotor: Dr. P. Mohajerin Esfahani

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr. P. Palensky	Delft University of Technology, promotor
Dr. P. Mohajerin Esfahani	Delft University of Technology, copromotor

Independent members:

Prof.dr.ir. J. van den Berg	Delft University of Technology
Prof.dr. X. Yu	Royal Melbourne Institute of Technology
Prof.dr.ir. J. A. la Poutré	Delft University of Technology
Dr. A. K. Srivastava	Washington State University
Dr.ir. J. L. Rueda Torres	Delft University of Technology
Prof.dr.ir. M. Zeman	Delft University of Technology, reserve member

This research was carried out with financial supports from the China Scholarship Council (Scholarship No. 201506020114).

Keywords: combined attacks, disruptive multivariate intrusions, vulnerability assessment, cyber risk analysis, robust attack detection

Printed by: Ipskamp

Cover design by Liping Chen

ISBN 978-94-028-1975-5

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

Dedicated to my country,
my parents, my love.

“愿以寸心寄华夏，且将岁月赠山河。”

NOTATION

$\mathbb{R}$	Set of real numbers
$\mathbb{N}$	Set of integer numbers
$\lvert\mathcal{V}\rvert$	Cardinality of set $\mathcal{V}$
$A \in \mathbb{R}^{m\times n}$	Real-valued matrix with m rows and n columns
$x \in \mathbb{R}^{n_x}$	Real-valued column vector of dimension n_x
I_m	Identity matrix of dimension m
$A(i,:)$	The i-th row of the matrix A
$x(i)$	The i-th entry of the vector x
t	Continuous-time instant, real-valued
k	Discrete-time instant, integer-valued
$x(t)$	Continuous-time vector variable
$x[k]$	Discrete-time vector variable
$\lVert x\rVert_p$	The p-norm of the vector x for $p \geq 1$
$x^T,\ A^T$	Transpose of vector x, matrix A
$\mathrm{Im}(A)$	The range space of matrix A
$\mathrm{diag}(x)$	Diagonal matrix with vector x sitting on the main diagonal
$\mathrm{diag}\big[A_1, \ldots, A_n\big]$	Block matrix with main diagonal elements of the matrices $A_1, \ldots, A_n$

List of Abbreviations

ACE	Area control error

AGC	Automatic generation control
AI	Artificial intelligence
AVR	Automatic voltage regulator
BDD	Bad data detection
CA	Contingency analysis
CPS	Cyber-physical system
CUSUM	Cumulative sum
DAE	Differential-algebraic equation
DNN	Deep neural networks
DNP	Distributed network protocol
DoS	Denial-of-Service
EMS	Energy management system
FDI	False data injection
HIL	Hardware-in-the-loop
HLA	High level architecture
HMI	Human-machine interface
ICA	Independent component analysis
ICS	Industrial control systems
ICT	Information and communication technology
IoT	Internet-of-Things
LAN	Local-area network
LP	Linear program
MDD	Missing data detection
MILP	Mixed integer linear program
PCA	Principle component analysis

PDC	Phasor data concentrator
PDF	Probability distribution function
PMU	Phasor measurement unit
QP	Quadratic program
RTU	Remote terminal unit
SCADA	Supervisory control and data acquisition
SC-OPF	Security constrained optimal power flow
SE	State estimation
VPN	Virtual private network
WAN	Wide-area network
WLS	Weighted least squares

CONTENTS

SUMMARY

The digital transformation of power systems has introduced a new challenge for robustness: cyber security threats. The recent cyber incidents against power systems, such as the Stuxnet worm attack and the hacker-caused Ukraine blackout, do illustrate the features of a potent attack that can have extensive resources to corrupt multiple data channels by both integrity and availability, and also the strong capability to keep stealthy from possible detectors. The majority of research has focused on pure data integrity or availability attacks from a specific aspect of vulnerability or impact assessment. However, vulnerability or even cyber risk analysis methods for combined data integrity and availability attacks are, lacking and in need to be developed. Besides, the current detection mechanisms of power systems are mainly for erroneous data and thus may fail in the presence of coordinated data corruptions. This thesis contributes to vulnerability and cyber risk analysis of power systems to combined attacks, and robust attack detection approaches.

First, a vulnerability assessment framework for power systems under combined attacks is developed. A concept of security index is introduced to calculate the attack resources needed by an attacker who may have full or limited knowledge of the targeted system. Here the full knowledge assumption which is commonly used in the literature is relaxed. Power systems are considered more vulnerable to attacks with smaller security index since such attacks can be executed with fewer resources. The detection probability of the combined attack with limited knowledge is also computed, which is a necessary step to derive its likelihood. After considering the attack impact on load estimates, a cyber risk metric is proposed to quantify the likelihood and impact of each attack in a comprehensive way. It is shown that combined attacks can bring higher risk to power system operations in most considered cases, comparing with other pure types of attacks.

Second, the analytic vulnerability assessment framework is extended to incorporate power system communication network properties and a co-simulation plat-

form is developed for cyber security analysis. The two parts of work on analytic assessment and numerical simulation in a lot of research are conducted independently, and this thesis aims to close this gap. The network properties such as the communication topology and the data routing scheme are modeled in the security index formulation. It is shown that power systems are still more vulnerable to combined attacks and multi-path routing can be adopted for attack mitigation. Besides, a co-simulation platform is built to capture the character of a cyber-physical power system, along with a thorough discussion on the coupling of power system and ICT infrastructure simulators. The attack scenarios from the analytic results are used to perform simulations for vulnerability validation and impact evaluation.

The third contribution is to develop a detector called diagnosis filter to reveal the occurrence of a type of disruptive data integrity attacks which may stay stealthy from the current bad data detection mechanism. Unlike some existing work, this thesis goes beyond a static viewpoint of attack detection to capture the attack impact on the dynamics of system trajectories. The diagnosis filter approach is then characterized as robust optimization programs where two possible desired features are investigated: (i) a non-zero transient and (ii) a non-zero steady-state behavior of the filter in the presence of attacks. Linear programming relaxation for the resulting robust program of (i) and even convex reformulations for (ii) are proposed, which improves the scalability, and as such practicality, of the filter design. The results of the latter yield a Nash equilibrium between the attack and the diagnosis filter, which implies that this detector is not based on a conservative design.

In the end, this thesis aims to implement the developed model-based diagnosis filter in a real or simulated power system. A further robustification scheme to minimize the effects from possible model mismatches on the filter output is developed, with the assistance of simulation data. This has contributed to a novel data-assisted model-based attack detection approach. The model mismatch signatures are extracted and an optimization-based framework is built to robustify the diagnosis filter to the model mismatches. Besides, the thesis also presents an approach on how to isolate each attack and even track the attack magnitude in the steady-state behavior of the filter. The effectiveness of the theoretical results is validated by simulations on an IEEE benchmark system in DIgSILENT PowerFactory.

SAMENVATTING

De digitale transformatie van energiesystemen heeft een nieuwe uitdaging voor robuustheid geïntroduceerd: cyberveiligheidsbedreigingen. De recente cyberincidenten tegen elektriciteitssystemen, zoals de Stuxnet-wormaanval en de door hackers veroorzaakte black-out in Oekraïne, illustreren de kenmerken van een krachtige aanval die uitgebreide eigenschappen kan hebben om meerdere datakanalen te beschadigen door zowel integriteit als beschikbaarheid, en ook de mogelijkheid om zich verborgen te houden van mogelijke detectoren. Het merendeel van het onderzoek is gericht op pure gegevensintegriteit of beschikbaarheidsaanvallen vanuit een specifiek aspect van kwetsbaarheids- of effectbeoordeling. Kwetsbaarheids- of zelfs cyberrisicoanalysemethoden voor gecombineerde gegevensintegriteit en beschikbaarheidsaanvallen ontbreken echter en moeten worden ontwikkeld. Daarbij zijn de huidige detectiemechanismen van elektriciteitssystemen hoofdzakelijk voor foutieve gegevens en kunnen dus falen in de aanwezigheid van gecoördineerde gegevenscorrupties. Dit proefschrift draagt bij aan kwetsbaarheids- en cyberrisicoanalyse van elektriciteitssystemen voor gecombineerde aanvallen en robuuste benaderingsdetectiemethoden.

Eerst wordt een kwetsbaarheidsbeoordelingskader ontwikkeld voor elektriciteitssystemen onder gecombineerde aanvallen. Een concept van beveiligingsindex wordt geïntroduceerd om de aanvalsbronnen te berekenen die een aanvaller nodig heeft die volledige of beperkte kennis van het beoogde systeem heeft. Hier is de veronderstelling van volledige kennis die in de literatuur veel wordt gebruikt, versoepeld. Elektriciteitssystemen worden als kwetsbaarder beschouwd voor aanvallen met een kleinere beveiligingsindex, omdat dergelijke aanvallen met minder middelen kunnen worden uitgevoerd. De detectiekans van de gecombineerde aanval met beperkte kennis wordt ook berekend, wat een noodzakelijke stap is om de waarschijnlijkheid hiervan af te leiden. Na de impact van de aanval op schattingen van de belasting beschouwd te hebben, wordt een cyberrisicometriek voorgesteld om de

waarschijnlijkheid en impact van elke aanval op een uitgebreide manier te kwantificeren. Het is aangetoond dat gecombineerde aanvallen in de meest overwogen gevallen een hoger risico kunnen opleveren voor de werking van elektriciteitssystemen, vergeleken met andere pure soorten aanvallen.

Ten tweede is het analytische kwetsbaarheidsbeoordelingskader uitgebreid met eigenschappen van het communicatiesysteem van het elektriciteitssysteem en is een co-simulatieplatform ontwikkeld voor analyse van cyberveiligheid. De twee delen van het werk over analytische beoordeling en numerieke simulatie in veel ander onderzoek worden onafhankelijk uitgevoerd en dit proefschrift beoogt deze kloof te dichten. De netwerkeigenschappen zoals de communicatietopologie en het gegevensrouteringsschema worden gemodelleerd in de beveiligingsindexformulering. Het is aangetoond dat elektriciteitssystemen nog kwetsbaarder zijn voor gecombineerde aanvallen en multi-path routing kan worden aangenomen voor aanvalsbestrijding. Bovendien is een co-simulatieplatform gebouwd om het karakter van een cyber-fysiek elektriciteitssysteem vast te leggen, samen met een grondige discussie over de koppeling van elektriciteitssysteem- en ICT-infrastructuursimulatoren. De aanvalsscenario's van de analyseresultaten worden gebruikt om simulaties uit te voeren voor validatie van kwetsbaarheden en impactevaluatie.

De derde bijdrage is het ontwikkelen van een detector die diagnosefilter wordt genoemd om het optreden van een soort verstorende gegevensintegriteitsaanvallen te onthullen die verborgen kunnen blijven voor het huidige mechanisme voor detectie van foutieve gegevens. In tegenstelling tot bestaand werk gaat dit proefschrift verder dan een statisch gezichtspunt van aanvalsdetectie om de aanvalsimpact op de dynamiek van systeemtrajecten vast te leggen. De diagnosefilterbenadering wordt vervolgens gekenmerkt als robuuste optimalisatieprogramma's waarbij twee mogelijke gewenste functies worden onderzocht: (i) een niet-nul transiënt en (ii) een niet-nul steady-state gedrag van het filter in de aanwezigheid van aanvallen. Lineaire programmaversnelling voor het resulterende robuuste programma van (i) en zelfs convexe herformuleringen voor (ii) worden voorgesteld, hetgeen de schaalbaarheid en als zodanig praktisch van het filterontwerp verbetert. De resultaten van deze laatste leveren een Nash-evenwicht op tussen de aanval en het diagnosefilter, wat inhoudt dat deze detector niet gebaseerd is op een conservatief ontwerp.

Uiteindelijk wil dit proefschrift het ontwikkelde modelgebaseerde diagnosefil-

ter implementeren in een echt of gesimuleerd elektriciteitssysteem. Met behulp van simulatiegegevens is een verder robuustheidsschema ontwikkeld om de effecten van mogelijke mismatches op de filteruitvoer te minimaliseren. Dit heeft bijgedragen aan een nieuwe, op gegevens gebaseerde, modelgebaseerde aanpak voor het detecteren van aanvallen. De model-mismatch-handtekeningen worden geëxtraheerd en een op optimalisatie gebaseerd raamwerk is gebouwd om het diagnosefilter te robuust maken voor de model-mismatches. Bovendien presenteert het proefschrift ook een benadering voor het isoleren van elke aanval en zelfs het volgen van de aanvalsomvang in het steady-state gedrag van het filter. De effectiviteit van de theoretische resultaten wordt gevalideerd door simulaties op een IEEE-benchmarksysteem in DIgSILENT PowerFactory.

1

INTRODUCTION

This chapter provides an introduction to this thesis. Motivations along with most relevant adversarial examples lead to two parts of the thesis work: vulnerability analysis and attack detection for intelligent power systems under complex cyber attacks. For each part, the research questions are given, and the contributions from each chapter are presented. Finally, the structure of this thesis is outlined.

1

1.1. MOTIVATIONS AND RESEARCH QUESTIONS

The increasingly digitized power system offers more data, details and controls in a real-time fashion than its non-networked predecessors. One of the benefiting applications of this development is the Energy Management System (EMS): Remote sensors provide measurement data via Information and Communication Technology (ICT) infrastructure such as Supervisory Control and Data Acquisition (SCADA) system. This measurement information is then used and processed by the EMS in a SCADA control center for State Estimation (SE), Automatic Generation Control (AGC), and decision making, etc. The security of energy supply depends on the EMS, which in turn depends on a reliable SCADA network.

However, vulnerabilities within ICT infrastructure have made the power system exposed to cyber security threats. SCADA systems, which are notorious for being based on legacy ICT, are a popular target for adversaries [1, 2]. Most SCADA network protocols, e.g., Modbus (Plus) Protocol, Distributed Network Protocol (DNP3), IEC 60870-5 and IEC 61850, are not designed to provide robust security checks at the time of publishing [3]. Besides, SCADA systems are more connected to corporate and Internet networks, leading to an increased number of vulnerabilities for malicious cyber adversaries to exploit. Figure 1.1 depicts an overview of the SCADA network and possible cyber security threats. Intrusions can originate from the corporate or Internet network (I1, I2), or the control center network (I3), or the (neighbor) substation network (I4, I5), or the remote access points (I6) to the targeted facilities. Once an intruder gains access to the SCADA network, he can disrupt the time synchronization of all protocols, compromise the availability of communications, or even control or modify the data or settings of the sensors and actuators [4]. Notably, the risks posed to power system SCADA networks are far greater in terms of the impact and scale of attacks than common computer security ones. Attacks on SCADA systems can result in poor situation awareness and incorrect system operations, affecting the power system reliability and economy aspects, or even causing cascading outages [5, 6].

Motivational examples. There have been a lot of real-world examples of SCADA disruptions by cyber attackers. In the following, two most relevant examples are used to illustrate the cyber risks from deliberate malware and adversaries.

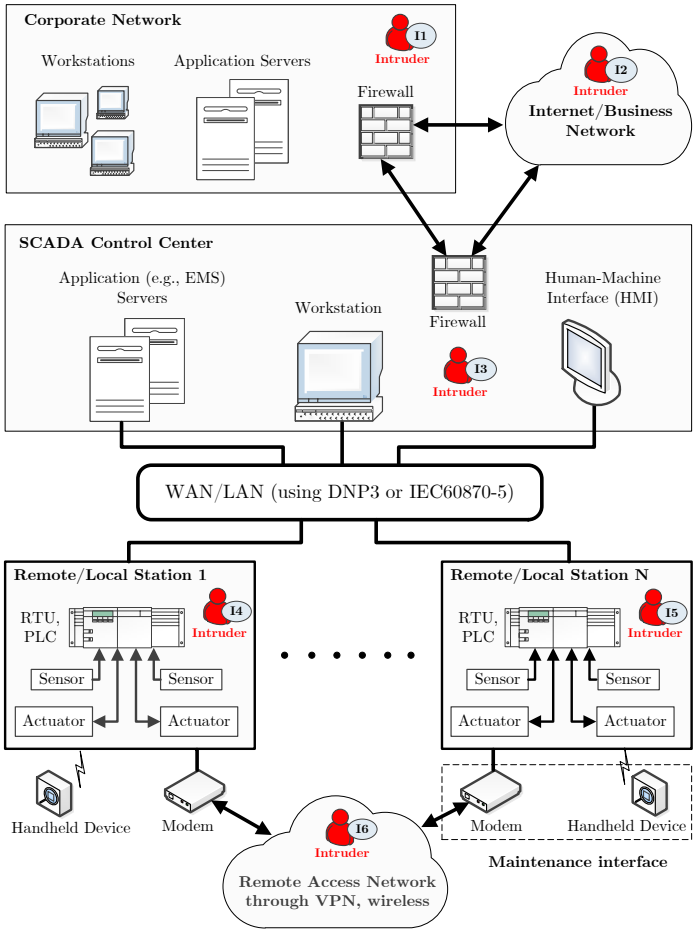


Figure 1.1: Overview of the SCADA network and cyber security threats based on [7, Figure 3]. A SCADA system is typically made up of a control center (or master) and remote substations. The SCADA center contains application (e.g., EMS) servers, workstations and a human-machine interface (HMI) that collects data or information from remote/local stations and sends back control commands through wide-area or local-area networks (WAN/LAN). A remote/local station consists of various types of critical components, e.g., sensors and control devices wired to the programmable logic controller (PLC) or directly interfaced with the remote terminal unit (RTU). The SCADA system is connected to the corporate or Internet networks through firewalls. Besides, substations can be remotely accessed via a virtual private network (VPN) or wireless communications for monitoring and maintaining. Cyber intrusions can originate from outside or inside the SCADA system where possible locations of intruders are illustrated.

Example 1: *Stuxnet* worm attack. Among all the diverse malware, *Stuxnet* is the most complex and functional one that aims at Industrial Control Systems (ICS) such as SCADA networks of critical power grid infrastructures. The major characteristics of the *Stuxnet* worm attack include [8, 9],

- extremely selective targeting - from the vulnerable PCs to program PLCs;
- four zero-day exploits - an unusually high number;
- remarkably detailed knowledge of PLCs and ICS;
- using a Windows rootkit to prevent a PC owner from discovering;
- high effort level and huge amount of attack resources;
- great lengths to remain dormant and strong capabilities of self-updating;
- a large number of infected hosts and organizations.

As we can see, *Stuxnet* worm attack has an unexpected combination of advanced skills, inside system knowledge, vast attack resources, great ability to keep stealthy. The malware contains codes for a deliberate attack that can fake the sensor measurements and control signals in the SCADA system without being detected.

Example 2: first hacker-caused 2015 Ukraine blackout. It is believed that the *Stuxnet* worm attack has infected numerous PCs and caused damages to almost a thousand of industrial facilities [10]. More recently, the first hacker-caused blackout happened in Ukraine on 23 December 2015. This cyber attacks entailed several technical components [11],

- a long-term reconnaissance of the victim networks to learn the environment and system knowledge;
- a prior compromise of corporate networks by BlackEnergy3 malware via spear phishing emails;
- a hijack of the SCADA network, remotely switching substations off;
- a disruption on the SCADA infrastructures, e.g., RTUs, modems;
- a destruction of master boot records in servers and workstations with the modified KillDisk firmware;
- a telephonic denial-of-service attack to jeopardize outage reports.

The most key feature in this event is that the attackers can perform long-term reconnaissance operations required to learn the system knowledge and execute a highly synchronized, multi-site attack [12]. These aspects, along with the complexity and functionalities of *Stuxnet*, do contribute to the feasibility of a potent SCADA network attack that it can be equipped with extensive system knowledge, enough attack resources to manipulate multiple sensors or actuators (i.e., multivariate attacks) in a coordinated manner and strong capability of remaining stealthy from

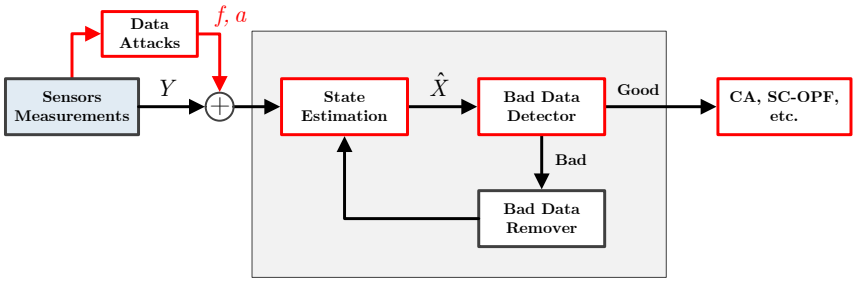


Figure 1.2: The SE process under data attacks. Y and $\hat{X}$ denote the measurements collected in sensors of substations and the estimated states of the power network. Besides, f and a represent false data injection (FDI) attack and data availability attack (e.g., Denial-of-Service (DoS) attack), respectively. CA: Contingency Analysis; SC-OPF: Security Constrained Optimal Power Flow.

possible detectors, etc. Thus this thesis is motivated to assess the vulnerability and cyber risk of intelligent power systems to such type of "smart" attacks, and come up with robust defense actions to reveal their occurrence.

1.1.1. PART I: VULNERABILITY ANALYSIS OF POWER SYSTEMS TO ATTACKS

A typical class of data integrity attacks can carefully launch synthesized false data injections (FDI) on a number of SCADA measurements. This multivariate attack can bypass the bad data detection (BDD) mechanism within the EMS without triggering alarms. The adversary here is able to (i) have full knowledge of the system model (e.g., topology information and system model parameters); (ii) manipulate multiple measurements with enough attack resources; (iii) keep stealthy from the detection schemes and achieve specific targets. These capabilities do capture the features of current cyber attacks against power system SCADA networks, as discussed in the adversarial examples above.

It was first explored in [13] that such a *stealthy multivariate attack* can perturb the SE function of the EMS without being detected. Figure 1.2 shows the major control loop of the SE process. Since state estimates are inputs of many application-specific tools in the EMS, the corrupted estimates can infect further control actions. Considering that SE is based on the power flow model, vulnerability analysis of power systems in steady-state to stealthy multivariate attacks has been a prominent subject in the literature. This vulnerability of power systems to stealthy attacks is usually quantified by computing the attack resources needed by the adversary to

alter specific measurements and keep stealthy, with full knowledge of the system model [14–16]. For that purpose, a concept of *security index* is usually introduced with a formulation of the following optimization program,

$$\begin{aligned} \alpha^\star := \quad & \min_{f} \quad \|f\|_p \\ & \text{s.t.} \quad f \in \mathcal{G}\,,\ f \in \mathcal{S}\,, \end{aligned} \tag{1.1}$$

where $\alpha^\star$ is the so-called security index, f denotes the FDI attack. The constraints in (1.1) for the attacks are scenario specific. Suppose that the attacker wishes to derive an attack vector f satisfying a set of goals (encoded by $f \in \mathcal{G}$) and remaining stealthy from possible detectors (i.e., $f \in \mathcal{S}$). The objective function in the sense of p-norm characterizes different metrics for least "attack effort": when $p = 0$, it denotes the minimum number of measurements to be corrupted, and the program (1.1) becomes non-convex; when $p = 1$, it may be used as a convex relaxation for the case of $p = 0$; when $p = 2$, it is related to the measurement redundancy of the system [17]. For each p, this proxy metric assesses "how hard" it is for the adversary to attack the specific system, and it is of interest to both the EMS operator and the attacker: if $\alpha^\star$ is large, it requires significantly coordinated attack resources by the adversary to accomplish; if $\alpha^\star$ is small, some of the measurements are critical as they require fewer corruptions to be altered stealthily. Hence, power systems here are considered more vulnerable to attacks with smaller security index.

The stealthy multivariate attack described above still needs intensive attack resources such as the capability to corrupt the integrity on a number of measurements. Denial-of-service (DoS) attacks [18, 19], a type of availability attack, are much "cheaper" to achieve, especially if sensors communicate via insecure communication channels. As shown in Figure 1.2, availability attacks can also take place on the EMS together with FDI attacks. Besides, notably, most of the work in the literature still assumes that the adversaries have full knowledge of the system, while in practice, an attacker may acquire a perturbed system model as a result of analyzing an out-dated or estimated model [20, 21]. Intuitively, an adversary can launch data availability attacks to block the measurements that he has the least knowledge of in the system. To this end, vulnerability analysis should also involve attack impact in the notion of *cyber risk* assessment. Thus firstly, this thesis focuses on answering the following group of questions that still remain insufficiently answered:

Q1 Vulnerability assessment: *How to assess the vulnerability of power systems to stealthy multivariate attacks? How about the vulnerability of the power system when it comes to combined attacks that both data integrity and availability attacks are launched? How combined attacks can differ from pure FDI multivariate attacks when both of them have incomplete knowledge of the system or limited attack resources? What would be the total cyber risks that the combined attacks can bring to the power system?*

In order to increase the security of EMS, one first needs analytic methods above to assess the vulnerability and cyber risk of power systems to attacks and then uses appropriate tools to validate and explore the attack scenarios. Some tools based on co-simulation techniques that integrate simulated power systems, communication networks and controls have been developed to analyze the behavior of intelligent power grids including cyber security issues [22–24]. However, these two parts of work are usually conducted independently even though they are related. Analytic methods may have to ignore some details when modeling the intelligent power grids, but could be used to guide the cyber security experiments on co-simulation tools, while the tools can support the cyber security analysis with empirical results. This could contribute to develop more robust algorithms or methods that combine system-theoretic and ICT-specific measures to protect the EMS against data attacks [25]. The first part of this thesis also aims to close this gap by answering the following research questions,

Q2 Co-simulation for cyber security analysis: *What tools can be adopted to aid in the cyber security analsis of intelligent power grids which have been given the character of a cyber-physical system (CPS)? How to couple hybrid power system and ICT simulators for co-simulation of the intelligent power grids? How to extend the developed vulnerability analysis framework for answering* **Q1** *to incorporate communication network properties and develop a co-simulation platform to conduct cyber security analysis?*

1.1.2. PART II: FROM STATIC TO DYNAMIC AND ROBUST DETECTION

National Institute of Standards and Technology (NIST) [26] defines five functions for protecting ICT infrastructure: (i) Identify, (ii) Protect, (iii) Detect, (iv) Respond,

(v) Recover. An ICT system can never be sufficiently protected without solving the problems of (iii-v). Besides, according to the bowtie model and the cyber risk management cycle in [27], to reduce impacts of an cyber incident, repressive measures need to be taken like the measures related to detection and recovery. Thus, the second part of this thesis work focuses on (iii) detection of multivariate data injection attacks on the SCADA system. It is of vital importance to detect cyber attacks and respond in an appropriate manner, as attacks on SCADA systems may bring disastrous economic and societal consequences. If the stealthy multivariate attacks in Section 1.1.1 can be detected in time, the corrupted signals can be disconnected or corrected by resilient controls, preventing further severe damages [28].

As mentioned above, SCADA systems deploy the detector BDD to filter out possible erroneous measurements due to sensor failures or anomalies [17, 29]. At each time step, the BDD process is performed to compute the measurement residual and check if it is below some threshold. The BDD mechanism is essentially a *static* detection scheme because it only captures a snapshot of the steady-state system trajectories. Although this method can detect some basic attacks, it may fail in the presence of *stealthy multivariate attacks* that launch carefully synthesized data injections. Detection methods have been proposed to reveal such stealthy attacks. Statistical methods, such as sequential detection using Cumulative Sum (CUSUM)-type algorithms were designed and discussed in [30]. In recent work [31, 32], anomaly detectors leverage additional information such as load forecasts, generation schedules and secure phasor measurement unit (PMU) data to generate diagnosis signal. These methods are, however, essentially *static* detection approaches that may be limited by some assumptions that measurements and states fit specific distributions or parts of the sensors are secure, while absolute cyber security is unattainable. Despite an extensive and ongoing literature focusing on the static part of the BDD mechanism, the following question remains largely unexplored:

Q3 From static to dynamic detection: *Would it be possible to detect stealthy multivariate attacks in a real-time operation by exploiting the attack impact on the dynamics of system trajectories during the transient behavior? If yes, how the robustness of the diagnosis tool can be ensured that the detector keeps sensitive to plausible disruptive multivariate attacks?*

The challenge of answering the above question and designing a robust diag-

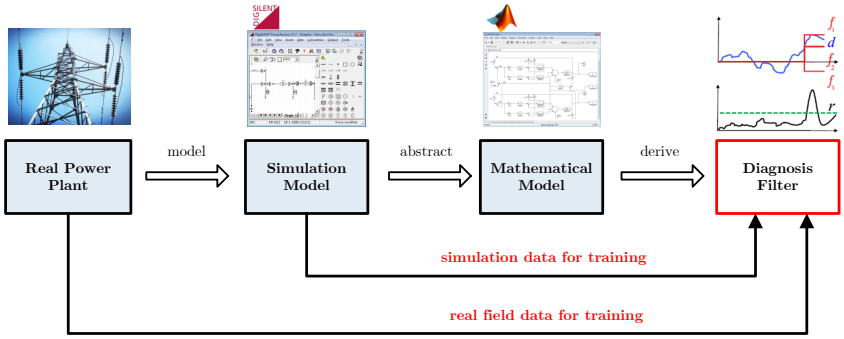


Figure 1.3: The process of data-assisted model-based anomaly detection. f_1, f_2, f_3 are signals of elements in a multivariate attack $f = [f_1 \ f_2 \ f_3]^\top$, d denotes unknown disturbance and r is a diagnosis signal (e.g., residual) for detection.

nosis tool is that, stealthy multivariate attacks may neutralize the detector outputs - cancel each other out due to multiple attacked signals. To overcome that, unlike the static detection schemes, the second part of this thesis work aims to design a robust diagnosis tool with a perspective of *dynamic residual generation*. For this purpose, the impact of stealthy attacks on the system dynamics needs to be characterized through a modeling framework (e.g., a set of differential equations). The robustness of the residual generator is achieved when some quantities are satisfied. For instance, a function $\mathcal{J}(\delta, \alpha)$ can be defined to reflect these quantities which are influenced by both the action $\delta \in \mathcal{D}$ of the detector (the dynamic residual generator) and the action $\alpha \in \mathcal{A}$ of the attacker. $\mathcal{D}$ and $\mathcal{A}$ are sets for describing the detector's and attacker's actions respectively. A successful scenario from the perspective of a powerful attacker may be to minimize this function given the knowledge of the diagnosis tool (i.e., δ). Therefore, we can take a rather conservative viewpoint where the attacker may have knowledge about the diagnosis tool and exploits it so as to synthesize a stealthy attack. Then the diagnosis tool design can be formulated as a robust optimization program,

$$\gamma^\star := \max_{\delta \in \mathcal{D}} \min_{\alpha \in \mathcal{A}} \Big\{ \mathcal{J}(\delta, \alpha) \Big\},$$

where $\gamma^\star$ is the optimal value that reflects the robustness of the diagnosis tool: if the obtained $\gamma^\star$ is larger than a certain value, it offers a robust residual generator that detects the plausible multivariate attacks in an admissible set.

The diagnosis tool design above can be classified as a *model-based* anomaly detection approach, which utilizes the explicit model of system dynamics to de-

tect stealthy multivariate attacks. Another approach referred to *data-driven* method tries to automatically learn the system characteristics from available data [33, 34]. In general, each type of these two methods has its own advantages and limitations. The effectiveness of model-based method depends on the "accuracy" of the model of power system dynamics. However, the models generated by complex power systems are mostly high-dimensional and nonlinear. What makes things worse is that an accurate model of a power system in real-time is always inaccessible. The second part of this thesis work is also motivated to improve the developed diagnosis tool towards real implementations by overcoming these challenges.

Consider the *plant-model mismatch* depicted in Figure 1.3 of power systems. When the full model of a whole power plant may be unavailable, high-fidelity simulators (e.g., DIgSILENT PowerFactory) are always used to describe the detailed power system to provide greater insights into its behavior. The simulation model in simulators can be a detailed numerical model, while the mathematical model characterizing the physical laws in the form of dynamical systems or differential equations can be simplified [1] to enhance its applicability (maybe at the cost of its effectiveness) in the model-based detection. Indeed, there exist mismatches between the power plant, its simulation model and its mathematical model. It becomes nontrivial to implement a model-based diagnosis tool in a real or detailed simulated power system as it may encounter such *model mismatches*. In this regard, this thesis aims to improve the model-based detector to be implemented in a high-fidelity simulator. The idea is to extract the mismatch signatures between the simulation model and the mathematical model with the assistance of simulation data. Based on these signatures, the diagnosis tool can be "trained" to be robustified to possible model mismatches. Then the robustfied detector can be "tested" on revealing the presence of attacks in the simulator. This would also bridge the gap between these two types of model-based and data-driven approaches, and make a step forward to a real implementation of the developed diagnosis tool, resulting in a novel *data-assisted model-based* approach.

Q4 Robust attack detection: *How to robustify the model-based diagnosis tool against possible model mismatches with the assistance of data from high-fidelity*

[1] For instance, the load frequency model of Automatic Generation Control (AGC) can be linearized and decoupled from the voltage dynamics.

simulators? Can these methods be implemented in a real or detailed simulated power system? Can these methods be further improved to always trigger alerts in the presence of multivariate attacks and even isolate or identify them?

1.2. CONTRIBUTIONS AND THESIS OUTLINE

In Part I of this thesis, Chapter 2 and 3 aim to provide a thorough analysis of multivariate data injection attacks and even combined attacks against intelligent power grids, i.e., the level of system knowledge and attack resources required by an adversary, and contribute to develop a co-simulation tool for supporting such an analysis. Next in Part II, Chapter 4 and 5 aim to develop a diagnosis tool to reveal the stealthy multivariate attacks with a scalable and robust design and latent capacity of implementation in the real-world power systems. The contributions consist of theoretical results, numerical simulations, tools design and developments, and are summarized as follows:

- **Chapter 2.** First it is shown that in theory, the optimal solution of the combined attacks security index problem coincides with the one of the pure FDI multivariate attacks security index problem. Chapter 2 continues to tackle **Q1** that the detection probability of the BDD mechanism under combined attacks with limited adversarial knowledge is computed. It is also shown that the optimal combined attack with limited knowledge can still keep stealthy under certain conditions. To this end, a cyber risk metric is proposed for the combined attacks with limited knowledge. Based on the analysis of risk metrics of combined attacks and FDI multivariate attacks, it is found that power system operations face higher risk under combined attacks.

- **Chapter 3.** It contributes to extend the vulnerability assessment framework of Chapter 2 to incorporate communication network properties and developing a co-simulation platform to conduct simulations for cyber security analysis, answering **Q2**. The communication network properties such as topology and routing schemes are modeled in the analytic vulnerability assessment framework. The fundamentals and coupling issues in co-simulations of intelligent power grids are also presented, along with modeling and simulation challenges. Additionally, experiments of the attack scenarios from the vulner-

ability analysis are conducted on the developed co-simulation platform.

- **Chapter 4.** To address **Q3**, Chapter 4 develops a diagnosis filter to detect the stealthy multivariate attacks in a real-time operation. Unlike some existing work based on a static viewpoint of detection, this chapter captures the attack impact on the dynamics of system trajectories. A dynamic residual generator approach is introduced and formulated as robust optimization programs. Besides, two desired features are investigated: (i) a non-zero transient and (ii) a non-zero steady-state behavior of the residual generator. For (i), a linear programming relaxation which is highly tractable for large-scale systems is proposed; for (ii), it is found that an exact convex reformulation and a Nash equilibrium between the attacker and the detector do exist. The latter implies that the proposed approach is not conservative as even the additional information of the attack signal does not improve the diagnosis performance.

- **Chapter 5.** The work of this chapter moves one step further to robustify the model-based diagnosis tool in Chapter 4 towards a real implementation in the power system operations, w.r.t. **Q4**. The concept of "model mismatch" is introduced to illustrate the possible difference between the detailed simulation model in a simulator (e.g., DIgSILENT PowerFactory) and the (simplified) mathematical model based on which the diagnosis filter is developed in Chapter 4. The patterns of model mismatches are extracted from the simulation data and then a novel data-assisted model-based approach is developed for robust attack detection. It is illustrated that the filter residual remains sensitive to multivariate attacks and even can isolate and track the attack value in the steady-state behavior, while keeping the effects from possible model mismatches on the residual output minimized.

The contributions above are based on published or submitted articles during my PhD study. In the following, a number of publications are listed according to the related parts of the thesis.

Part I: Vulnerability Analysis

K. Pan, A. Teixeira, M. Cvetkovic, & P. Palensky (2018). Cyber Risk Analysis of

Combined Data Attacks Against Power System State Estimation. *IEEE Transactions on Smart Grid*, 10(3), 3044–3056. DOI: 10.1109/TSG.2018.2817387;

P. Palensky, A. van der Meer, C. López, A. Joseph, & **K. Pan** (2017). Applied Cosimulation of Intelligent Power Systems: Implementing Hybrid Simulators for Complex Power Systems. *IEEE Industrial Electronics Magazine*, 11(2), 6–21. DOI: 10.1109/MIE.2017.2671198;

P. Palensky, A. van der Meer, C. López, A. Joseph, & **K. Pan** (2017). Cosimulation of Intelligent Power Systems: Fundamentals, Software Architecture, Numerics, and Coupling. *IEEE Industrial Electronics Magazine*, 11(1), 34–50. DOI: 10.1109/MIE.2016.2639825;

K. Pan, A. Teixeira, C. López, & P. Palensky (2017). Co-simulation for Cyber Security Analysis: Data Attacks against Energy Management System. In *8th IEEE International Conference on Smart Grid Communications*, Dresden, Germany, 253-258. DOI: 10.1109/SmartGridComm.2017.8340668;

K. Pan, A. Teixeira, M. Cvetkovic, & P. Palensky (2017). Data Attacks on Power System State Estimation: Limited Adversarial Knowledge vs. Limited Attack Resources. In *43rd Annual Conference of the IEEE Industrial Electronics Society*, Beijing, China, 4313-4318. DOI: 10.1109/IECON.2017.8216741;

M. Cvetkovic, **K. Pan**, C. David López, R. Bhandia, & P. Palensky (2017). Co-simulation Aspects for Energy Systems with High Penetration of Distributed Energy Resources. In *2017 AEIT International Annual Conference*, Cagliari, Italy, 1-6. DOI: 10.23919/AEIT.2017.8240488;

K. Pan, A. Teixeira, M. Cvetkovic, & P. Palensky (2016). Combined Data Integrity and Availability Attacks on State Estimation in Cyber-Physical Power Grids. In *7th IEEE International Conference on Smart Grid Communications*, Sydney, Australia, 1-7. DOI: 10.1109/SmartGridComm.2016.7778773.

Part II: Attack Detection

K. Pan, P. Palensky, & P. Mohajerin Esfahani (2019). From Static to Dynamic Anomaly Detection with Application to Power System Cyber Security. *IEEE Transactions on Power Systems*, pp, 1–1. DOI: 10.1109/TPWRS.2019.2943304;

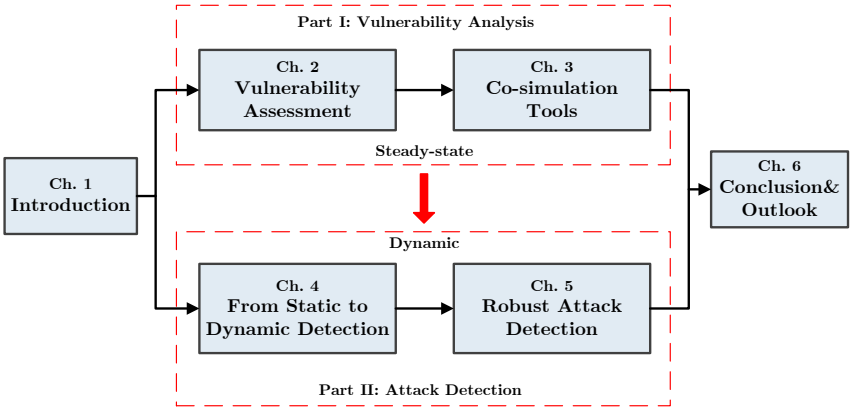


Figure 1.4: Outline of This thesis.

K. Pan, P. Palensky, & P. Mohajerin Esfahani (2019). Robust Attack Detection in Smart Grids: A Novel Data-assisted Model-based Approach. To be submitted to *IEEE Transactions on Power Systems*;

K. Pan, D. Gusain, & P. Palensky (2019). Modelica-Supported Attack Impact Evaluation in Cyber Physical Energy System. In *IEEE 19th International Symposium on High Assurance Systems Engineering*, Hangzhou, China, 228–233. DOI: 10.1109/HASE.2019.00042.

The research work of this thesis follows a natural flow as illustrated in Figure 1.4. The outline is as follows. Chapter 2 conducts theoretical vulnerability assessment of power systems in steady-state to stealthy multivariate attacks and combined attacks. In Chapter 3, the analytic vulnerability assessment framework of Chapter 2 is extended to incorporate communication network properties for a better characterization of cyber-physical systems. A developed co-simulation tool for supporting cyber security analysis is also presented. Chapter 4 moves from the analysis of power system in steady-state to system dynamics that a diagnosis tool approach is developed to detect the class of stealthy multivariate attacks by exploiting the attack impact on the dynamics of system trajectories. An improvement of the diagnosis tool is introduced in Chapter 5 in which it is further robustified to possible model mismatches with the assistance of simulation data from high fidelity simulators. Conclusions are drawn in Chapter 6 where recommendations for future research are also provided.

2

VULNERABILITY ANALYSIS OF POWER SYSTEMS IN STEADY-STATE TO DATA ATTACKS

Understanding smart grid cyber attacks is key for developing appropriate protection and recovery measures. Advanced attacks pursue maximized impact at minimized costs and detectability. This chapter conducts vulnerability assessment of combined data integrity and availability attacks against the power system state estimation. The combined attacks are compared with pure false data injection (FDI) attacks - multivariate attacks. A security index for vulnerability assessment to these two kinds of attacks is proposed and formulated as a mixed integer linear program. It is shown that such combined attacks can succeed with fewer resources than FDI multivariate attacks. The combined attacks with limited knowledge of the system model also ex-

This chapter is based on the following published work:
[35] K. Pan, A. Teixeira, M. Cvetkovic, & P. Palensky (2018). Cyber Risk Analysis of Combined Data Attacks Against Power System State Estimation. *IEEE Transactions on Smart Grid*, 10(3), 3044–3056. DOI: 10.1109/TSG.2018.2817387;
[36] K. Pan, A. Teixeira, M. Cvetkovic, & P. Palensky (2016). Combined Data Integrity and Availability Attacks on State Estimation in Cyber-Physical Power Grids. In *7th IEEE International Conference on Smart Grid Communications*, Sydney, Australia, 1-7. DOI: 10.1109/SmartGridComm.2016.7778773.

pose advantages in keeping stealthy from the bad data detection mechanism. Finally, the risk of combined attacks to reliable system operation is evaluated using the results from vulnerability assessment and attack impact analysis. The findings in this chapter are validated and supported by a detailed case study.

2.1. INTRODUCTION

The State Estimation (SE) within modern energy management systems (EMS) is an instance of the dependency between the physical power system and the ICT infrastructures. It provides the operator with an estimate of the system state with the inputs of power flow measurements delivered by the SCADA system. Nowadays, SE has been an integral tool in EMS for contingency analysis (CA), security-constrained optimal power flow (SC-OPF), and pricing calculation algorithms, etc. The critical nature of SE highlights the importance of making it accurate and secure for power system operations. However, As discussed in Section 1.1.1, the SCADA system is vulnerable to a large number of security threats. False data injection attacks, as a typical class of data integrity attack, have been studied with considerable attention. By modifying a number of measurements coordinately, this multivariate attack can pass the BDD mechanism within the SE to stay stealthy [13] from the operators.

2.1.1. STATE-OF-THE-ART

Research in the literature has focused on FDI attacks against the SE from many aspects of cyber risk assessment [37], e.g., vulnerability analysis, attack impact evaluation and mitigation schemes development. As first shown in [13], a class of FDI attack, so-called *stealthy multivariate attack*, can perturb the state estimate without triggering alarms in the BDD scheme within the SCADA networks. Vulnerability of SE to stealthy multivariate attacks is quantified by computing attack resources needed by the attacker to manipulate specific measurements and keep stealthy from the BDD scheme, with or without full knowledge of the system model [14–16]. Note that for a broader review of vulnerability analysis, in a lot of research work, the power system structural vulnerability to failures or intentional attacks is also quantified using complex network techniques [38]. Method based on topological models has been a major subclass, among which the work in [39] first proposed maximum-flow-based approach to access line vulnerability with a new centrality index, and a hybrid model taking into account both the complex network and the power flow characteristics was developed in [40]. A cascading faults graph approach considering both topological and operational vulnerabilities can be found in [41].

As shown in Figure 2.1, SE provides inputs for other applications in EMS and if it got corrupted, it can infect further control actions, misleading operators to dis-

ruptive decisions. The estimate errors due to stealthy multivariate attacks were analyzed in [42] and [43]. The results illustrate that the errors could be significant even with a small number of measurements being compromised. The work in [6] and [44] studied the potential economic impact of multivariate attacks against the SE by observing the nodal price of market operation. The attacker could obtain economic gain or cause operating costs in the market. Recent work in [45] studied the physical impact of such attacks with the attacker's goal to cause a line overflow.

In order to defend against stealthy multivariate attacks, mitigation schemes have been proposed to improve the bad data detection algorithm or safeguard certain measurements from adversarial data injection. Sequential detection (or quickest detection) of stealthy attacks was designed mainly based on well-known Cumulative Sum (CUSUM) algorithm in [30]. In reference [46], detection methods that leverage synchrophasor data and other forecast information were presented. The network layer and application layer mitigation schemes, such as multi-path routing and data authentication and protection, are proved to be effective to decrease the vulnerability of power systems to these attacks [47] [36].

Most of the research above assumes that the adversary has full knowledge of the system model including the power network topology and parameters. However, the data of the system model is usually key protected and the attacks are always executed with limited adversarial knowledge. The work in [48, 49] proposed that an FDI attack can be made with incomplete network information. The attacker can still keep stealthy if it knows the local information (topology and transmission line parameters) of the attacking region under certain conditions. The authors also explored how to launch a successful FDI attack against AC state estimation with incomplete knowledge [50]. Another limited knowledge scenario is that the attacker has inaccurate network information of the power system [20]. Such FDI attacks have the probability to be detected by the BDD mechanism while the detectability is intimately related to the detectability of topology or parameter errors [17]. For these limited knowledge cases, the adversary could also infer necessary network information based on available data using learning methods such as independent component analysis (ICA) [51] and subspace estimation technique [52].

It is worth noting that the majority of research has focused on stealthy multivariate attacks from a specific aspect of vulnerability or impact assessment. The

work in [19] first considered adding a class of availability attacks, so-called jamming attack, to the attack scenarios against the SE. Our paper [36] first studied the *stealthy combined attacks* with different measurement routing topologies, concluding that such attacks may need less attack resources than FDI multivariate attacks. Besides, the work above still assumed that the attackers have perfect knowledge of the system model. In practice, we are more interested in the limited adversarial knowledge case that the attacker knows inaccurate network information. Such attacks are not guaranteed to be stealthy. This chapter would like to explore how combined attack can differ from multivariate attacks in a limited knowledge setting. Intuitively, combined attacks provide the availability attack option to block measurements that the attacker has least knowledge of. This motivates the use of attack resources and the detection probability attacks with limited knowledge in vulnerability assessment. In addition, vulnerability and impact of attacks can be combined together in the notion of *cyber risk*. In [53], a high-level risk assessment methodology for power system applications including SE was presented. However, risk analysis methods and tools combining vulnerability and impact assessment for data attacks are needed to implement risk assessment methodologies.

In this chapter, for the first time the combined attacks with limited knowledge of the system model are formulated and the vulnerability analysis of combined attacks is conducted. To do that, it first analyzes the vulnerability of SE with respect to attack resources needed by the attacker and calculates the detection probability of combined attacks with limited system knowledge. This is a necessary step in deriving the likelihood of the attack. Next, an impact metric is proposed for evaluating attack impact on load estimate. Combining the results from vulnerability and impact assessment, the risk which attacks bring to reliable system operations is presented. This chapter compares the vulnerability, impact and risk with those of FDI multivariate attacks. The simulation results show that combined attacks yield higher risk in majority of considered cases.

2.1.2. CONTRIBUTIONS AND OUTLINE

To the best of my knowledge, this chapter work is the first one to conduct vulnerability analysis of combined attacks with limited knowledge of the system model. The contributions are listed as follows:

(i) The first part of vulnerability analysis is presented through the notion of *security index* [15], which corresponds to the minimum attack resources needed by the attacker to compromise the measurements while keeping stealthy (programs (2.10) and (2.11)). The power system is more vulnerable to attacks with smaller security index since such attacks can be executed with fewer resources. It is shown that, the optimal solution of combined attack security index problem coincides with the optimal solution of the FDI multivariate attack security index problem (Theorem 2.3.3 and Corollary 2.3.4).

(ii) The second contribution is to address the detection probability problem of combined attacks with limited adversarial knowledge. Here the full knowledge assumption which is commonly adopted in the work of literature is relaxed. It is shown that the optimal combined attack with limited adversarial knowledge can still keep stealthy from the current BDD mechanism under certain conditions (Theorem 2.4.2 and Proposition 2.4.3). The empirical results also indicate that combined attacks can have lower detection probability.

(iii) A cyber risk metric is proposed to quantify the risk of combined attacks with limited knowledge of the system. For the attacks with the same security index, the risk metric is computed by multiplying (i) the probability of the attack not to be detected, with (ii) the attack impact on load estimate (Algorithm 1). The attack impact on load estimate is particularly considered because such estimates are inputs of other applications that compute optimal control actions in EMS (Definition 2.5.1). Based on the analysis of risk metrics of combined attacks and FDI attacks, it is shown that power system operations face higher risk under combined attacks scenario.

The outline of this chapter is as follows. Section 2.2 gives an introduction of SE and stealthy multivariate attacks mechanism. Section 2.3 extends the attack scenario to combined attacks and proposes security index with computational method for vulnerability analysis. In Section 2.4, the detectability of combined attacks with limited adversarial knowledge is discussed. The risk metric is proposed to measure the risk of attacks in Section 2.5 with the analysis of the vulnerability and attack impact. Section 2.6 presents empirical results from a power system use case.

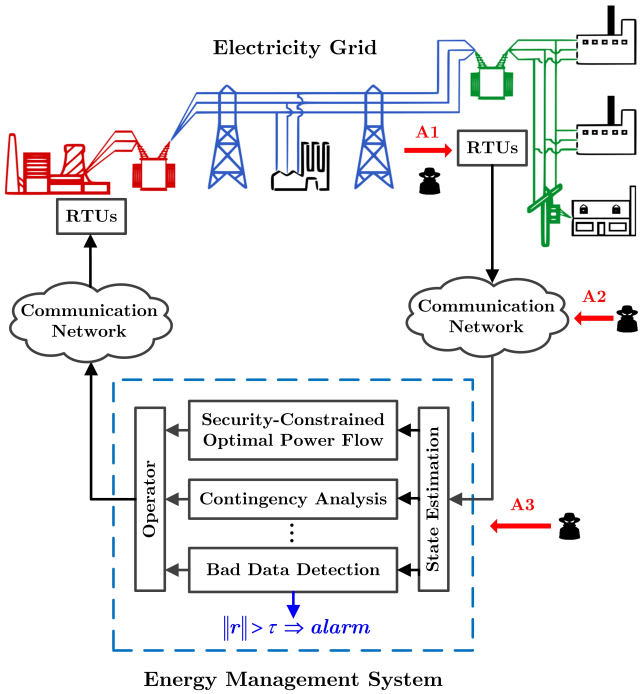


Figure 2.1: Schematic diagram of the electricity grid, SCADA communication network and EMS based on [54, 55]. The SE function uses power flow measurements collected by RTUs and transmitted through the SCADA system to estimate the current state of the power system. An alarm is triggered by the bad data detection when the norm of the residual signal r exceeds a given threshold τ.

2.2. SYSTEM MODELING AND STEALTHY ATTACKS

State estimation uses measurements collected by the remote terminal units and transmitted through the SCADA communication network to estimate the current state of the system. There is a built-in BDD mechanism to detect erroneous measurements. The estimated state is then processed by other application specific tools such as the CA and SC-OPF modules to compute optimal control actions while ensuring reliability and safety. Figure 2.1 depicts the whole closed-loop system process. Cyber attacks can manipulate the measurements by directly tampering the RTUs (A1) in substations, the SCADA network (A2), or even the databases and application servers in the control center (A3). In this section, the system modeling approach and the BDD techniques are reviewed. Besides, the stealthy multivariate attacks problem is introduced.

2.2.1. STATE ESTIMATION

The considered power system has $n_b + 1$ buses and n_t transmission lines. Considering the power system in steady-state (power flow model), the data collected by RTUs includes line power flow and bus power injection measurements. These n_Y measurements are denoted by $Y = [Y_1, \dots, Y_{n_Y}]^\top$. The system state X is the vector of phase angles and voltage magnitudes at all buses except the reference bus whose phase angle is set to be zero. For the analysis of cyber security and bad data detection in SE, it is customary to describe the dependencies of measurements and system state through an approximate model called DC power flow model [16]. In the simplified DC power flows, all the voltage magnitudes are assumed to be constant and the reactive power is completely neglected. Thus the vector Y refers to active power flow and injection measurements, and the state X refers to bus phase angles only. There are n_X phase angles to be estimated excluding the reference one, i.e. $X = [X_1, \dots, X_{n_X}]^\top$. Hence, Y and X are related by the equation

$$Y = P \begin{bmatrix} WB^T \\ -WB^T \\ B_0 WB^T \end{bmatrix} X + e := Cx + e, \tag{2.1}$$

where $e \sim \mathcal{N}(0, R)$ is the measurement noise vector of independent zero-mean Gaussian variables with the covariance matrix $R = \text{diag}(\sigma_1^2, \dots, \sigma_{n_Y}^2)$, $C \in \mathbb{R}^{n_Y \times n_X}$ represents the system model, depending on the topology of the power network, the line parameters and the placement of RTUs. Here the topology is described by a directed incidence matrix $B_0 \in \mathbb{R}^{(n_X+1)\times n_t}$ in which the directions of the lines can be arbitrarily specified [16]. Matrix $B \in \mathbb{R}^{n_X \times n_t}$ is the truncated incidence matrix with the row in B_0 corresponding to the reference bus removed. The line parameters are described by a diagonal matrix $W \in \mathbb{R}^{n_t \times n_t}$ with diagonal entries being the reciprocals of transmission line reactance. Matrix $P \in \mathbb{R}^{n_Y \times (2n_t + n_X + 1)}$ is a matrix stacked by the rows of identity matrices, indicating which power flows or bus injections are measured. Usually a large degree of measurements redundancy is employed to make C full rank.

The state estimate $\hat{X}$ can be obtained by the following weighted least squares (WLS) estimate:

$$\hat{X} := \arg\min_X (Y - CX)^\top R^{-1} (Y - CX), \tag{2.2}$$

which can be solved as $\hat{X} = (C^\top R^{-1} C)^{-1} C^\top R^{-1} Y := KY$. The estimated state $\hat{X}$ can

be used to estimate the active power flows and injections by

$$\hat{Y} = C\hat{X} = CKY := TY, \tag{2.3}$$

where T is the so-called hat matrix [29]. The BDD scheme uses such estimated measurements to identify bad data by comparing $\hat{Y}$ with Y, see below.

2.2.2. A STATIC DETECTOR - BAD DATA DETECTION

Measurements data may be corrupted by random sensor errors. Thus there is a built-in BDD module in SCADA networks for the purpose of bad data detection. The BDD mechanism is achieved by hypothesis tests using the statistical properties of the measurement *residual*:

$$r = Y - \hat{Y} = (I - T)Y := SY = Se, \tag{2.4}$$

where $r \in \mathbb{R}^{n_Y}$ is the residual vector, $I \in \mathbb{R}^{m\times m}$ is an identity matrix and S is the so-called residual sensitivity matrix [29]. It can be seen that the BDD is a static detector which concerns only a snapshot of the steady-state system trajectories.

This chapter mainly considers the $J(\hat{X})$-test based BDD technique. For the measurement error $e \sim \mathcal{N}(0, R)$, the new random variable $\varepsilon = \sum_{i}^{n_Y} R_{ii}^{-1} e_i^2$ where R_{ii} is the diagonal entry of the covariance matrix R has a χ^2 distribution with $n_Y - n_X$ degrees of freedom. Note the quadratic cost function $J(\hat{X}) = \|R^{-1/2}r\|_2^2 = \|R^{-1/2}Se\|_2^2$. fFor the independent n_Y measurements we can have $\text{rank}(S) = n_Y - n_X$, which implies that $J(\hat{X})$ has a so-called *generalized chi-squared distribution* with $n_Y - n_X$ degrees of freedom [56]. The BDD mechanism uses the quadratic function as an approximation of ε and checks if it follows the distribution χ^2_{m-n}. Defining $\alpha \in [0, 1]$ as the significance level corresponding to the false alarm rate, and $\tau(\alpha)$ such that

$$\int_0^{\tau(\alpha)} f(x)dx = 1 - \alpha, \tag{2.5}$$

where $f(x)$ is the probability distribution function (PDF) of χ^2_{m-n}. Hence, the BDD scheme becomes

$$\begin{cases} \text{Good data, if} & \|R^{-1/2}r\|_2 \le \sqrt{\tau(\alpha)}, \\ \text{Bad data, if} & \|R^{-1/2}r\|_2 > \sqrt{\tau(\alpha)}. \end{cases} \tag{2.6}$$

2.2.3. STEALTHY MULTIVARIATE ATTACKS

The goal of an attacker is to perturb the SE while keeping stealthy from the BDD. If only data integrity attacks are considered, the attacker could inject false data on a set of measurements, modifying the vector Y into $Y_f := Y + f$ where the *multivariate attack vector* $f \in \mathbb{R}^{n_Y}$ represents the false data injections. A k_f-tuple multivariate false data injection attack is defined as follows,

Definition 2.2.1 (k_f-tuple FDI multivariate attack)**.** *An FDI attack with an attack vector $f \in \mathbb{R}^{n_Y}$ is called a k_f-tuple FDI multivariate attack if a number of k_f measurements are injected with false data, i.e. $\|f\|_0 = k_f$ where $\|f\|_0$ denotes the number of non-zero elements in the vector f.*

As shown in [13], an attacker with full knowledge of the system model (i.e., the matrix C in the algebraic equation (2.1)) and the capability to corrupt a specific number of measurements can keep stealthy if it follows $f = C\Delta X$ where $\Delta X \in \mathbb{R}^{n_X}$ is non-zero. The corrupted measurements Y_f becomes $Y_f = C(X + \Delta X) + e$. This leads to the state estimate perturbed by a degree of ΔX, while the residual for BDD checking remains unchanged. It has been verified that such *stealthy multivariate attacks* based on the DC model can be performed on a real SCADA/EMS testbed with full nonlinear AC system model while avoiding the bad data detection scheme [42].

To describe the vulnerability of the SE to these stealthy multivariate attacks, a *security index* is introduced to compute the minimum number of measurements to be corrupted by the attacker to keep stealthy [15]. The security index is given by

$$\begin{aligned} \alpha_j^\star := \quad & \min_{\Delta X} \quad \|f\|_0 \\ & \text{s.t.} \quad f = C\Delta X,\ f(j) = \mu, \\ & \qquad\ f(l) = 0,\ \forall l \in \mathcal{L}, \end{aligned} \tag{2.7}$$

where $f(j)$ denotes the injected false data on measurement j, and μ is the non-zero *attack magnitude* determined by the attacker. The second constraint is added that the pseudo-measurements corresponding to zero-injection buses and full protected me-asurements (all in the set $\mathcal{L}$) cannot be attacked. The result $\alpha_j^\star$ is the security index that quantifies the vulnerability of measurement j to stealthy multivariate attacks. Here the computed $\alpha_j^\star$ belongs to one of the multivariate attacks with the minimum k_f ($k_f = \alpha_j^\star$) for measurement j. It is known that the optimization problem above is NP-hard (See [57]). In [16], the authors proposed an approach using

the big M method to directly express (2.7) as a mixed integer linear programming (MILP) problem which can be solved with an appropriate solver,

$$\begin{aligned} \alpha_j^\star = \min_{\Delta X,\, w} \quad & \sum_{i=1}^{n_Y} w(i) \\ \text{s.t.} \quad & C\Delta X \le Mw, \\ & -C\Delta X \le Mw, \\ & C(j,:)\Delta X = \mu, \\ & C(l,:)\Delta X = 0,\ \forall l \in \mathcal{L}, \\ & w(i) \in \{0,1\},\ \forall i. \end{aligned} \tag{2.8}$$

In (2.8), M is a constant scalar that is greater than the maximum absolute value of entries in $C\Delta X^\star$, for some optimal solution $\Delta X^\star$ of (2.7). At optimality, for any i that $|C(i,:)\Delta X^\star| = 0$, the corresponding $w(i)$ is zero. Thus an optimal solution to (2.8) is exactly the same optimal solution to (2.7) with $w(i) = 1$ indicating that the measurement i is corrupted by an FDI attack. Here the *attack magnitude* μ is determined by the attacker and is set as a tunable parameter in the optimization problem (2.8). Thus, the attacker can vary the attack magnitude based on the possible constraints arising from the presence of measurement forecasts and range limitations. Then the optimization problem (2.8) which computes FDI multivariate attacks is denoted as $\mathbf{P}_f(C)$ where C corresponds to the full system model.

2.3. VULNERABILITY ANALYSIS FOR COMBINED ATTACKS

Stealthy multivariate attacks are resource-intensive since the adversary needs to coordinate integrity attacks on all targeted measurements. This usually gives the adversary more power than possible in practice [43]. In reality, an attacker would try to reduce the attack resources and would prefer data availability attacks (e.g., DoS attacks, jamming attacks) since SCADA systems are always more vulnerable to this types [58]. Thus this section focuses on the adversarial scenario where the attacker would launch combined data integrity and availability attacks.

2.3.1. COMBINED DATA INTEGRITY AND AVAILABILITY ATTACKS

For a large-scale SCADA system, missing data and failing RTUs are common [15]. When some of the measurements are missing, the typical solution widely employed in SE is to use the remaining data before the system becomes “unobservable”. An-

other solution is to use pseudo-measurements (e.g., previous data, forecast information), but these measurements would still lose confidence in further time intervals as long as the availability attacks continue. The combined attacks introduced here are attacks which will not make the system unobservable or lead to non-convergence of the SE algorithm. It can be said that such combined attacks can still keep stealthy from the detector BDD, with the following definition.

Definition 2.3.1 (stealth combined attacks)**.** *Attacks which can launch both availability attack and FDI attack are called stealthy combined attacks if no additional alerts are triggered in the current BDD mechanism.*

In practice, the current BDD mechanism within SE would not trigger alarms when some measurements are missing. Besides, even when availability attacks happen, they may be misdiagnosed as poor network conditions or physical damages to the sensors. Thus the assumption is kept in this chapter that the SE uses remaining data if availability attacks take place and they would not trigger additional alerts in the BDD. Let $a \in \{0,1\}^{n_Y}$ be the *availability attack vector* for the availability attacks and $a(i) = 1$ means that measurement i is unavailable. Thus the model for remaining measurements and system state can be described by

$$Y_a = C_a x + e_a, \tag{2.9}$$

where $Y_a \in \mathbb{R}^{n_Y}$ and $e_a \in \mathbb{R}^{n_Y}$ are the measurement and noise vectors respectively, and the entries of them are zero if the corresponding measurements are unavailable. Matrix $C_a \in \mathbb{R}^{n_Y \times n_X}$ denotes the model of the remaining measurements and it is obtained from C by replacing some rows with zero row vectors due to availability attacks on these measurements, i.e. $C_a := (I_{n^Y} - \mathrm{diag}(a))C$. Thus the hat matrix and residual sensitivity matrix when availability attacks occur can be obtained,

$$K_a := (C_a^\top R^{-1} C_a)^{-1} C_a^\top R^{-1},$$

$$T_a := C_a K_a, \quad S_a := I_{n^Y} - T_a.$$

For the combined attacks, the attacker would still launch FDI attacks on the remaining measurements in concert with availability attacks, making Y_a changed into $Y_{f,a} := Y_a + f$. Similarly, a (k_f, k_a)-tuple combined attack can be defined as

Definition 2.3.2 ((k_f, k_a)-tuple combined attack)**.** *A combined attack with an FDI attack vector $f \in \mathbb{R}^{n_Y}$ and an availability attack vector $a \in \{0,1\}^{n_Y}$ described above is called a (k_f, k_a)-tuple combined attack if $\|f\|_0 = k_f$, $\|a\|_0 = k_a$.*

2.3.2. SECURITY INDEX FOR COMBINED ATTACKS

Similar to the FDI multivariate attacks, if the attack vectors of a (k_f, k_a)-tuple attack satisfy $f = C_a \Delta X$, such combined attacks can still keep stealthy as the attack vector f lies on the range space of the matrix C_a. Using the formulation of security index in (2.7) for stealthy multivariate attacks, an intuitive security index for combined data integrity and availability attacks can be proposed as follows,

$$\begin{aligned} \beta_j^\star := \min_{\Delta X, a} \quad & \|f\|_0 + \|a\|_0 \\ \text{s.t.} \quad & f = C_a \Delta X, \\ & C_a = (I_{n^Y} - \text{diag}(a))C, \\ & f(j) = \mu, \\ & f(l) = 0, \ \forall l \in \mathcal{L}, \\ & a(k) \in \{0,1\}, \ \forall k. \end{aligned} \tag{2.10}$$

The result $\beta_j^\star$ is the security index that quantifies how vulnerable measurement j is to combined attacks. The computed $\beta_j^\star$ belongs to one of the combined attacks that have minimum $k_f + k_a$ ($k_f + k_a = \beta_j^\star$) for measurement j. To solve this NP-hard problem above, here a computational solution which uses the big M method to formulate (2.10) as a MILP problem is proposed:

Theorem 2.3.3. *For any measurement index $j \in \{1, \ldots, n_Y\}$ and non-zero attack magnitude μ, the optimization (2.10) can be equivalently described via the following MILP optimization program*

$$\begin{aligned} \beta_j^\star = \min_{\Delta X, w, a} \quad & \sum_{i=1}^{n_Y} w(i) + \sum_{k=1}^{n_Y} a(k) \\ \text{s.t.} \quad & C\Delta X \le M(w + a), \\ & -C\Delta X \le M(w + a), \\ & C(j,:)\Delta X = \mu, \\ & C(l,:)\Delta X = 0, \ \forall l \in \mathcal{L}, \\ & w(i) \in \{0,1\}, \ \forall i, \\ & a(k) \in \{0,1\}, \ \forall k, \end{aligned} \tag{2.11}$$

where $w, a \in \{0,1\}^{n_Y}$ *with* $w(i) = 1$ *and* $a(k) = 1$ *meaning FDI attack and data availability attack on measurements* i *and* k *respectively.*

Proof. The proof follows by re-writing (2.10) as (2.11). First, note that the constraint of (2.10), $f = (I_{n^Y} - \text{diag}(a))C\Delta X$, can be formulated as a set of inequality constraints with auxiliary binary variables by using the big M method, yielding $-Mw \leq (I_{n^Y} - \text{diag}(a))C\Delta X \leq Mw$, where $w \in \{0,1\}^{n_Y}$ and $\|f\|_0 = \sum w(i)$. Since a is a vector of binary variables, the pair of inequality constraints pertaining the i-th measurement can be written as $|(1-a(i))C(i,:)\Delta X| \leq Mw(i)$. The latter can be read as

$$\begin{cases} C(i,:)\Delta X = 0, & \text{if } w(i) = a(i) = 0, \\ |C(i,:)\Delta X| \leq M, & \text{if } w(i) = 1 \text{ or } a(i) = 1, \end{cases}$$

which can be rewritten as $|C(i,:)\Delta X| \leq M(a(i) + w(i))$. Hence, recalling that $f(i) = (1-a(i))C(i,:)\Delta X$, it can be concluded that the constraints of (2.10) can be equivalently rewritten as the constraints of (2.11). The proof concludes by noting that the objective functions of these two satisfy the equality $\|f\|_0 + \|a\|_0 = \sum w(i) + \sum a(i)$. ■

Similarly, the optimization program 2.11 which computes the combined attacks is denoted by $\mathbf{P}_{f,a}(C)$. By solving $\mathbf{P}_f(C)$ from (2.8) and $P_{f,a}(C)$ from (2.11), the system operators can obtain the attack vectors and further assess the risk of attacks on the measurements, which will be illustrated in Section 2.5.

Corollary 2.3.4. *For any measurement index* $j \in \{1, \ldots, n_Y\}$ *and non-zero attack magnitude* μ*, let* $(\Delta X^\star, w^\star, a^\star)$ *be an optimal solution to* (2.11)*. Then an optimal solution to* (2.7) *can be computed as* $\Delta X^\star$*, and* $\alpha_j^\star = \beta_j^\star$.

Proof. The proof follows straightforwardly from Theorem 2.3.3, which establishes that the optimization (2.10) can be equivalently described via the MILP program (2.11): comparing (2.11) and (2.8), it can be easily seen that an optimal solution to (2.8) can be computed as $(\Delta X^\star, w_a^\star)$ with $w_a^\star = w^\star + a^\star$, and $\alpha_j^\star = \beta_j^\star$. The proof concludes by noting that (2.8) is an exact MILP reformulation of (2.7). ■

Corollary 2.3.4 implies that a set of compromised measurements is an optimal solution to (2.10) if and only if this set is an optimal solution to (2.7), and the two security indexes $\beta_j^\star$ and $\alpha_j^\star$ coincide. In fact, in [59] it was shown that the set of compromised measurements in a k_f-tuple FDI multivariate attack obtained by solving

(2.7) is a sparsest *critical tuple* containing the target measurement j. A sparsest critical tuple is characterized by the measurements that do not belong to a critical tuple of lower order. A critical tuple contains a set of measurements, where removal all of them will cause the system to be unobservable. If any subset of the critical tuple is removed, it would not lead to the loss of observability [29]. According to Corollary 2.3.4 and its proof, the set of compromised measurements of multivariate attacks in this critical tuple is also an optimal solution to the security index problem (2.10) of combined attacks. The interpretation of the security index problem as a critical tuple problem provides the means for comparing security indexes of attacks with full and limited adversarial knowledge; see Section 2.4 for details.

The optimizations $\mathbf{P}_f(C)$ and $\mathbf{P}_{f,a}(C)$ derived so far in (2.8) and (2.11) could identify the compromised measurements set of attacks but did not consider the attack costs. In what follows, the costs are included in the formulation. To simplify the discussion, it is assumed that the availability and integrity attacks have the costs C_A and C_I, respectively, per measurement. Thus a security index for attack resources of combined attacks can be formulated, by rewriting the objective of (2.11) as

$$\theta_j^\star := \min_{\Delta X, w, a} \left\{ \sum_{i=1}^{n_Y} C_I w(i) + \sum_{k=1}^{n_Y} C_A a(k) \right\}, \tag{2.12}$$

where (2.12) has the same constraints as (2.11). It can be seen that the set of compromised measurements from the optimal solution of (2.12) is the optimal solution to (2.10) and also (2.7). If $C_A = C_I$, this is the same case as the one described in Corollary 2.3.4. For C_A and C_I with different values, see the following proposition.

Proposition 2.3.5. *When $C_A < C_I$, the optimal strategy of combined attack is to inject false data on the targeted measurement j and make other measurements in the critical tuple unavailable to the SE, yielding a $(1, \beta_j^\star - 1)$-tuple combined attack with optimal attack cost $\theta_j^\star = C_I + (\beta_j^\star - 1)C_A$. When $C_A > C_I$, the combined attack has the same optimal strategy as the FDI multivariate attack, i.e., injecting false data on the all measurements in the critical tuple, yielding a $(\beta_j^\star, 0)$-tuple combined attack (i.e., $\beta_j^\star$-tuple FDI attack) with optimal attack cost $\theta_j^\star = \beta_j^\star C_I$.*

Proof. If the values are taken that satisfy $C_A < C_I$, the optimal solution of $w^\star$ and $a^\star$ in (2.12), w.r.t. measurement j, would lead to $\sum w^\star(i) = 1$ and $\sum a^\star(k) = \beta_j^\star - 1$. This means that the optimal combined attack in the case of $C_A < C_I$ is to corrupt

one measurement with an integrity attack and make other measurements in this critical tuple unavailable. If the values are taken that satisfy $C_A > C_I$, the optimal solution of $w^\star$ and $a^\star$ in (2.12), w.r.t. measurement j, would lead to $\sum w^\star(i) = \beta_j^\star$ and $\sum a^\star(k) = 0$, i.e., the optimal combined attack is to inject false data on all the measurements in this critical tuple. ■

As previously indicated, availability attacks (e.g., DoS) can cost less attack resources compared with integrity attacks (e.g., FDI) in general. An intuitive example is that the attacker uses the same tool to perform a Man-In-The-Middle (MITM) attack on the exchanged measurements between substations and the SCADA control center. Thus the adversary is capable of interfering with the transmitted measurements using the MITM tool, either launching integrity or availability attacks. Unlike the integrity attack in which the attacker has to inject specific data values and repackage the data packets carefully, the availability attack only needs to block the measurements or modify them to zero or random errors [60]. Using the same MITM tool, the availability attacks become "cheaper" to achieve than integrity attacks. Of course, the true attack costs of different kinds of attacks launched by different tools are hard to quantify in practice. One possible way is to relate the attack cost to the inverse-likelihood of the attack. Likelihood assessment of attacks using attack trees or graphs also implies that availability attacks have higher probability to take place considering the factors of skills, knowledge and time [61]. Thus in the following of this chapter the values would satisfy $C_A \leq C_I$. The above Proposition 2.3.5 for the case $C_A < C_I$ will also be validated in Section 2.6.

2.4. ATTACKS WITH LIMITED ADVERSARIAL KNOWLEDGE

From this section the scenario in which the adversary has limited knowledge of the system model is considered and it is discussed how this affects the detectability of FDI multivariate attacks and combined attacks.

2.4.1. RELAXING ASSUMPTION ON ADVERSARIAL KNOWLEDGE

For the combined attacks and multivariate attacks above, the adversary is assumed to have full knowledge of system model C in (2.1) which consists of the network topology, the placement of RTUs and the transmission line reactance. This system

data is kept in the database of control center, and as such it is difficult to be accessed by the attacker [1]. This section extends the previous analysis by replacing the full knowledge assumption. Hence, in what follows the attacker only has limited knowledge of the system model, which may be more common in the real-world conditions. In particular, the limited knowledge case that is of interest to this thesis is the one that the attackers have inaccurate network information. The system model known by the adversary gets "perturbed" that system model uncertainties exist. An attacker could acquire perturbed system model as a result of analyzing an out-dated or estimated model using power network topology data but limited or inaccurate information of transmission line parameters [17, 20, 21].

From the attacker's perspective, without loss of generality, the perturbed system model known by the attacker can be denoted as $\tilde{C}$ which can be expressed as

$$\tilde{C} \triangleq C + \Delta C, \tag{2.13}$$

where $\Delta C \in \mathbb{R}^{n_Y \times n_X}$ denotes the part of model uncertainty. It is still assumed that the attacker uses the same linear policies to compute attack vectors, i.e. $f = \tilde{C}_a \Delta X$ for combined attacks and $f = \tilde{C} \Delta X$ for FDI multivariate attacks and $\tilde{C}_a := (I_{n^Y} - \text{diag}(a))\tilde{C}$. Correspondingly, the optimization program (2.7) in limited adversarial knowledge is denoted by $\mathbf{P}_f(\tilde{C})$ w.r.t $\tilde{C}$ computing FDI attacks and the optimization (2.10) is denoted by $\mathbf{P}_{f,a}(\tilde{C})$ w.r.t $\tilde{C}$ computing combined attacks.

2.4.2. DETECTABILITY OF ATTACKS WITH LIMITED KNOWLEDGE

Combined attacks When the measurements are corrupted by a (k_f, k_a)-tuple combined attack, the measurement residual $r_{f,d}$ can be written as

$$r_{f,a} = S_a Y_{f,a} = S_a e_a + S_a f. \tag{2.14}$$

As discussed in Section 2.3.2, when the attack vectors of the combined attack satisfy $f = C_a \Delta X$, the residual $r_{f,a} = S_a e_a + S_a C_a \Delta X = S_a e_a$ due to $S_a C_a = 0$, then the residual is not affected by f and no additional alarms are triggered; the BDD treats the measurements attacked by availability attacks as a case of missing data. However, for the attack with limited knowledge, the attack vector f becomes $f = \tilde{C}_a \Delta X$

[1] This can be an external attacker or an individual attacker, but an internal attacker or a state-sponsored attacker may have this capability to access such system knowledge or even more power to cause damages.

and $S_a f$ in (2.14) may be non-zero. In this case, the residual is incremented and the attack can be detected with some probability.

Note that the quadratic cost function with the combined attacks becomes $J_{f,a}(\hat{X}) = \|R^{-1/2}S_a e_a + R^{-1/2}S_a f\|_2^2$. Here the mean of $(R^{-1/2}S_a e_a + R^{-1/2}S_a f)$ is the non-zero $R^{-1/2}S_a f$ incremented by the attack. Recalling the $J(\hat{X})$-test based BDD, $J_{f,a}(\hat{X})$ has a *generalized non-central chi-squared distribution* with $n_Y - n_X - k_a$ degrees of freedom under the combined attacks. $J_{f,a}(\hat{X})$ can be used as an approximation of having the *non-central chi-squared distribution* $\chi^2_{n_Y-n_X-k_a}(\|R^{-1/2}S_a f\|_2^2)$ where the non-centrality parameter is $\|R^{-1/2}S_a f\|_2^2$ to calculate the detection probability $P_d(f, a)$ of combined attacks. Further such approximation would be validated using empirical results from Monte Carlo simulation in Section 2.6.2. It can be obtained that

$$\int_0^{\tau_{f,a}(\alpha)} f_{\lambda_{f,a}}(x)dx = 1 - P_d(f, a)\,, \tag{2.15}$$

where $f_{\lambda_{f,a}}(x)$ is the PDF of $\chi^2_{n_Y-n_X-k_a}(\|R^{-1/2}S_a f\|_2^2)$, $\tau_{f,a}(\alpha)$ is the threshold set in the BDD using (2.5) but with the PDF of $\chi^2_{n_Y-n_X-k_a}$.

FDI multivariate attacks For a k_f-tuple FDI attack with limited system knowledge, the quadratic function $J_f(\hat{X})$ can also be approximated to have a non-central chi-squared distribution but with $n_Y - n_X$ degrees of freedom. The distribution $\chi^2_{n_Y-n_X}(\|R^{-1/2}Sf\|_2^2)$ can be used to compute the detection probability $P_d(f)$ of FDI multivariate attacks. Similar to (2.15), $P_d(f)$ can be computed by solving

$$\int_0^{\tau(\alpha)} f_{\lambda_f}(x)dx = 1 - P_d(f), \tag{2.16}$$

where $\lambda_a = \|R^{-1/2}Sa\|_2^2$ denotes the non-centrality parameter, $\tau(\alpha)$ is the threshold set in the BDD using (2.5).

2.4.3. SPECIAL CASE: ATTACKS WITH STRUCTURED MODEL UNCERTAINTY

An interesting analysis is to understand what the model uncertainty ΔC in (2.13) is to the adversary. As stated in [17], the scenarios where the uncertainty is more structured are of greater interest. Here it is assumed that the attacker knows the exact topology of the power network, but has to estimate the line parameters. This assumption is feasible in the sense that the attacker can access the topology information by (i) collecting offline data such as topology maps and online data using

attacker's own meters; (ii) using market data to extract it from locational marginal prices; (iii) utilizing available power flow measurements and compromised breaker status data, as summarized in [62]. However, usually the attacker has limited access to the knowledge of the exact length of the transmission line and type of the conductor being used. Even if the attacker obtains such knowledge, the values would change by the time of implementing the attack due to weather conditions in temperature [20]. Denote the line parameters matrix with errors as $\tilde{W} \triangleq W + \Delta W$ where $\Delta W \in \mathbb{R}^{n_t \times n_t}$ represents the parameter uncertainty. Thus the model with this structured uncertainty becomes

$$\tilde{C} = P \begin{bmatrix} (W+\Delta W)B^T \\ -(W+\Delta W)B^T \\ B_0(W+\Delta W)B^T \end{bmatrix} \Rightarrow \Delta C = P \begin{bmatrix} \Delta W B^T \\ -\Delta W B^T \\ B_0 \Delta W B^T \end{bmatrix}. \tag{2.17}$$

Now let us consider the security index of attacks w.r.t. $\tilde{C}$ in (2.17). As discussed in Section III-B, the security index problem can be interpreted as a critical tuple problem. In the remaining part of this chapter the following assumption is adopted.

Assumption 2.4.1. *The system with perturbed model $\tilde{C}$ in (2.17) has the same sets of critical tuples as the system with original model C in (2.1).*

Assumption 2.4.1 is expected to hold in the case that the system with C in (2.1) is topologically observable [63]. Defining the security indexes for compromised measurements set under structured uncertainty model as $\tilde{\alpha}_j^\star$ and $\tilde{\beta}_j^\star$, respectively, the following theorem shows that the security index remains the same although the model is perturbed with structured uncertainty.

Theorem 2.4.2. *For any measurement index $j \in \{1, \ldots, n_Y\}$ and non-zero attack magnitude μ, under Assumption 2.4.1, let $(\Delta\tilde{X}^\star, \tilde{w}^\star, \tilde{a}^\star)$ be an optimal solution to $\mathbf{P}_{f,a}(\tilde{C})$ ($\tilde{C}$ is from (2.17)). Then there exists some $\Delta X^\star$ such that $(\Delta X^\star, w^\star, a^\star)$ with $w^\star = \tilde{w}^\star$ and $a^\star = \tilde{a}^\star$ is an optimal solution to $\mathbf{P}_{f,a}(C)$, $(\Delta X^\star, w_a^\star)$ with $w_a^\star = \tilde{w}^\star + \tilde{a}^\star$ is an optimal solution to $\mathbf{P}_f(C)$, and it satisfies $\tilde{\beta}_j^\star = \beta_j^\star = \alpha_j^\star = \tilde{\alpha}_j^\star$.*

Proof. The optimal solution with $\tilde{w}^\star$ and $\tilde{a}^\star$ identifies a sparsest critical tuple containing measurement j for the perturbed model $\tilde{C}$ in (2.17), which is also a sparsest critical tuple for the model C in (2.1) according to Assumption 2.4.1. Then the set of measurements in this critical tuple is an optimal solution to $\mathbf{P}_{f,a}(C)$. According to

Theorem 2.3.3 and Corollary 2.3.4, the set of measurements in this critical tuple is also an optimal solution to $\mathbf{P}_f(C)$. ■

With respect to the security index for attack resources in (2.12), let $\tilde{\theta}_j^*$ be the security index of attacks but w.r.t. perturbed model $\tilde{C}$. It can be seen that the set of compromised measurements from optimal solution to (2.12) w.r.t. $\tilde{C}$ in (2.17) is also the optimal solution to (2.11) and (2.8) according to Theorem 2.4.2. When it is the case that $C_A < C_I$, the optimal solution of $\tilde{w}^\star$ and $\tilde{a}^\star$ from (2.12) w.r.t. $\tilde{C}$, would lead to $\sum \tilde{w}^\star(i) = 1$ and $\sum \tilde{a}^\star(k) = \tilde{\beta}_j^\star - 1$. Such $(1, \tilde{\beta}_j^\star - 1)$-tuple combined attack can be launched with least attack resources when $C_A < C_I$ and the following shows that it can also achieve minimized detectability.

As discussed in Section IV-B, the detection probability would increase when attacker has limited knowledge of the system model. However, for the combined attacks, the following proposition states that the combined attacks with structured model uncertainty can still keep stealthy from the BDD if the following conditions are satisfied: (i) structured model uncertainty is defined as in (2.17); (ii) Assumption 2.4.1 holds.

Proposition 2.4.3. *For any measurement index $j \in \{1, \ldots, n_Y\}$ and non-zero attack magnitude μ, under Assumption 2.4.1, let $(\Delta\tilde{X}^\star, \tilde{w}^\star, \tilde{a}^\star)$ with $\sum \tilde{w}^\star(i) = 1$ be an optimal solution to $\mathbf{P}_{f,a}(\tilde{C})$ ($\tilde{C}$ is from (2.17)). Then this $(1, \tilde{\beta}_j^\star - 1)$-tuple combined attack from $(\Delta\tilde{X}^\star, \tilde{w}^\star, \tilde{a}^\star)$ is a stealthy attack.*

Proof. The FDI attack vector of this combined attack is $f = \tilde{C}_{\tilde{a}^\star}\Delta\tilde{X}$. According to Theorem 2.4.2, there exists $\Delta X^\star$ such that $(\Delta X^\star, w^\star, a^\star)$ with $w^\star = \tilde{w}^\star$ and $a^\star = \tilde{a}^\star$ is an optimal solution to $\mathbf{P}_{f,a}(C)$. Using the attack strategy above, $k_f = \sum \tilde{w}^\star(i) = 1$ and the only non-zero entry of the attack vector f is μ while other measurements in this critical tuple are attacked by availability attacks. Thus this combined attack is with the vector $f = (I_{n^Y} - \text{diag}(\tilde{a}^\star))\tilde{C}\Delta\tilde{X}^\star = (I_{n^Y} - \text{diag}(a^\star))C\Delta X^\star = C_{a^\star}\Delta X^\star$, which can keep stealthy w.r.t. C in (2.1). ■

It should be noted that, Proposition 2.4.3 is independent from the parameter uncertainty ΔW. This $(1, \tilde{\beta}_j^\star - 1)$-tuple combined attack can always keep stealthy for any parameter uncertainty levels as long as the critical tuple is correctly identified.

2.5. CYBER RISK METRICS FOR DATA ATTACKS

The previous sections focus on vulnerability assessment of SE to FDI multivariate attacks and combined attacks with limited knowledge. Following the procedure of cyber risk analysis in [53], this section defines and analyzes the *cyber risk* brought by attacks with limited knowledge. Usually the total *risk* of data attacks is defined as the likelihood of attack multiplied by the potential attack impact [37]. For a (k_f, k_a)-tuple combined attack, the risk metric $R(f,a)$ can be expressed as

$$R(f,a) = L(f,a) * I(f,a) \tag{2.18}$$

where $L(f,a)$ denotes the likelihood of the combined attack with attack vectors f and a, and $I(f,a)$ denotes the attack impact. For the attacks with larger risk metrics, they bring more risk to reliable system operation. The following sections discuss how to obtain $L(f,a)$ and $I(f,a)$.

2.5.1. LIKELIHOOD OF DATA ATTACKS

The attack likelihood relates to the vulnerability of the system. In this work, the likelihood of the attack is taken as the probability that the it is launched and the probability that it can keep stealthy from the detection mechanisms,

$$L(f,a) = P(f,a)P(s\,|\,f,a), \tag{2.19}$$

where $P(s\,|\,f,a)$ denotes the conditional probability of the attack passing the BDD if it has been performed successfully. For the attack with limited knowledge, the detection probability $P_d(f,a)$ can be obtained from Section 2.4.2 of (2.15). Thus it leads to $P(s|f,a) = 1 - P_d(f,a)$. In (2.19), $P(f,a)$ represents the probability that a particular adversary would perform a combined attack and successfully corrupt the data. Obtaining meaningful and realistic data for calculating $P(f,a)$ remains an unsolved and open issue for most of the established approaches [64]. The proposed security index $\tilde{\gamma}_j^\star$ w.r.t. perturbed model $\tilde{C}$ captures the efforts required by a combined attack and essentially can be related to the probability $P(f,a)$. It can be reasonably assumed that if the attacks have the same security index of $\tilde{\gamma}_j^\star$, they have the same probability of $P(f,a)$. In this chapter, to compare the risk of attacks with the same security index, $P(f,a)$ is "normalized" to be 1, meaning that the attacks have been performed successfully. The following risk metric applies to the combined attacks

with the same security index of $\tilde{\gamma}_j^\star$,

$$R(f,a) = P(f,a)P(s|f,a)I(f,a) = (1 - P_d(f,a))I(f,a). \tag{2.20}$$

For the k_f-tuple FDI attacks with the same security index of $\tilde{\gamma}_j^\star$, the formulation of risk metric is similar, i.e. $R(f) = (1 - P_d(f))I(f)$ where $I(f)$ denotes the attack impact and $R(f)$ is the risk metric. Thus with the results above, the risk of combined attacks and FDI multivariate attacks is comparable.

2.5.2. ATTACK IMPACT: ERRORS OF LOAD ESTIMATE

The estimated information from SE is used by other applications in EMS to compute optimal control actions. These are typically computed by minimizing network operation costs which are obtained by solving SC-OPF algorithms. As the work in [45] [65] shows, the SC-OPF application uses the load estimate from SE as the inputs. In practice, the important outputs from EMS are the injection estimate and SC-OPF results which would affect the further operations. If data attacks take place and pass the BDD, the load estimates get perturbed, which would influence control actions. Therefore, the impact metric can be formulated as a function of the bias introduced by the attack on the load estimate.

Assuming that the actual injections are described in a vector $G \in \mathbb{R}^{n_G}$ where n_G is the number of buses with injections, the impact on the errors of estimated power injections and actual power injections is considered,

$$\epsilon = \hat{G}_{f,a} - G, \tag{2.21}$$

where $\hat{G}_{f,a} \in \mathbb{R}^{n_G}$ is the vector of estimated injections under a (k_f, k_a)-tuple combined attack. Thus,

$$\epsilon = C_i \hat{X}_{f,a} - C_i X,$$

where $\hat{X}_{f,a} = K_a(Y_a + f) = X + K_a e_a + K_a f$, $C_i \in \mathbb{R}^{n_G \times n_X}$ denotes the submatrix of C by keeping the rows corresponding to injections including loads. It can be further obtained that $\epsilon = C_i K_a f + C_i K_a e_a$ where the term introduced by the attacks is $C_i K_a f$. Here K_a is the function of C_a as defined in (2.10). The expected value of ϵ is

$$\mathbb{E}(\epsilon) = C_i K_a f. \tag{2.22}$$

where $\mathbb{E}(\cdot)$ denotes the expectation. Thus the attack impact metric for combined attacks is defined as follows.

Definition 2.5.1. *The impact metric $I(f,a)$ for quantifying attack impact of a combined attack with FDI attack vector f and availability vector a on load estimate is defined as the 2-norm of $C_i K_a f$, i.e. $I(a,d) := \|C_i K_a f\|_2$.*

Similar to the combined attacks, the attack impact metric $I(f) = \|C_i K f\|_2$ is defined for a k_f-tuple FDI attack with attack vector f. The linear attack policy is still adopted to compute attack vectors for attacks with limited knowledge, i.e., $f = \tilde{C}_a \Delta X$ for combined attacks and $a = \tilde{C} \Delta X$ for FDI multivariate attacks.

Algorithm 1 Risk Assessment for Combined Attacks

1) Determine the attack magnitude μ. Compute attack vectors f and a from the optimization program $\mathbf{P}_{f,a}(\tilde{C})$.

2) Calculate the detection probability $P_d(f,a)$ of the combined attacks with f and a, according to the procedure in Section 2.4.2.

3) Calculate the attack impact metric $I(f,a)$ from Definition 2.5.1.

4) Compute the risk metric $R(f,a)$ for combined attacks by the formulation of (2.20) with the results from 2) and 3).

Giving all the information above, Algorithm 1 summarizes the risk assessment procedure for combined attacks and FDI attacks. First, the system operators would solve programs $\mathbf{P}_f(\tilde{C})$ and $\mathbf{P}_{f,a}(\tilde{C})$ w.r.t perturbed model $\tilde{C}$ for security indexes to compute the attack vectors. Then the detection probability of attacks and the attack impact could be obtained respectively according to Section 2.4.2 and Definition 2.5.1, leading to the risk metric of (2.20). Thus in conclusion, the risk assessment presented in this chapter, including the computation of attack vectors, the detection probability and the impact of attacks, provides insights at the planning stage of the power grid and offline analysis of combined attacks and FDI multivariate attacks in the limited knowledge case.

2.6. CASE STUDY

This section performs analysis on the IEEE benchmark system (Figure 2.2). The simulations are conducted on simplified DC power flow model for the purposes of: (i)

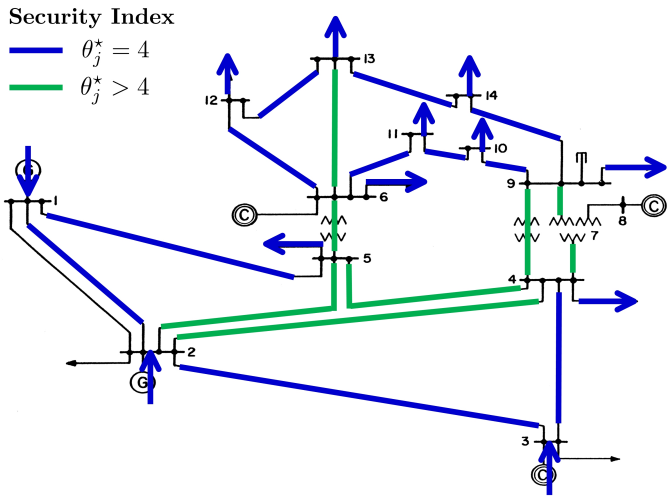


Figure 2.2: The IEEE 14-bus system. The measurements are labeled different colors according to their security index $\gamma_j^{a,d}$ from Figure 2.3. Here the vulnerable measurements with small index (= 4) are color coded blue. The measurements that have large index (> 4) are color coded green. The pseudo-measurements (without color) on bus 7, 8 and line 7-8 can not be attacked.

illustrating the vulnerability of power systems to combined attacks; (ii) providing insights into how combined attack can differ from FDI multivariate attack; (iii) evaluating the risk of data attacks and giving the risk prioritization. In the performed experiments, measurements are placed on all the buses and transmission lines to provide large redundancy; See Table 2.1. In the 14-bus system, measurements on bus 7, bus 8 and line 7-8 are pseudo-measurements for zero-injection buses and can not be attacked. The per-unit system is used and the power base is 100 MW. The measurements are generated with Gaussian noise ($\sigma_j = 0.02$ for measurement j). For the limited knowledge case, the attacks are under structured uncertainty model with the error on the line parameters of ±10%, ±20%, ±30% and ±40%.

2.6.1. SECURITY INDEX FOR VULNERABILITY ANALYSIS

In order to expose vulnerability of power systems to data attacks, the security index is calculated using the computation solutions of (2.11) (according to Theorem 2.3.3) and (2.8) for both combined attacks and FDI attacks. Thus the minimum number of compromised measurements and attack resources needed by the attacker to corrupt the SE process without being detected are determined. Figure 2.3 shows the

Table 2.1: The list of the 54 measurements on the branches and buses of the IEEE 14-bus system. P_{i-j} denotes the power flow measurement on the branch from Bus i to Bus j. P_i denotes the power injection measurement on Bus i.

1	P_{1-2}	13	P_{6-13}	25	P_{5-2}	37	P_{14-9}	49	P_9
2	P_{1-5}	14	P_{7-8}	26	P_{4-3}	38	P_{11-10}	50	P_{10}
3	P_{2-3}	15	P_{7-9}	27	P_{5-3}	39	P_{13-12}	51	P_{11}
4	P_{2-4}	16	P_{9-10}	28	P_{7-4}	40	P_{14-13}	52	P_{12}
5	P_{2-5}	17	P_{9-14}	29	P_{9-4}	41	P_1	53	P_{13}
6	P_{3-4}	18	P_{10-11}	30	P_{6-5}	42	P_2	54	P_{14}
7	P_{3-5}	19	P_{12-13}	31	P_{11-6}	43	P_3		
8	P_{4-7}	20	P_{13-14}	32	P_{12-6}	44	P_4		
9	P_{4-9}	21	P_{2-1}	33	P_{13-6}	45	P_5		
10	P_{5-6}	22	P_{5-1}	34	P_{8-7}	46	P_6		
11	P_{6-11}	23	P_{3-2}	35	P_{9-7}	47	P_7		
12	P_{6-12}	24	P_{4-2}	36	P_{10-9}	48	P_8		

security indexes $\theta_j^\star$ of attacks in the IEEE 14-bus system. Here the cost of FDI attack on per measurement is assumed to be 1 ($C_I = 1$) and $C_A = 0.5$ as $C_A/C_I = 0.5$ is considered. The x-axis indicates the measurement j targeted by the attacker with attack magnitude $\mu = 0.1$ p.u.. Note that in Figure 2.3 the pseudo-measurements 14, 34, 47, 48 from Bus 7, 8 and Branch 7-8 can not be attacked. The results illustrate the attack resources needed by the attacker to keep stealthy. The security index of combined attacks is also shown in Figure 2.2 where the measurements are color coded to indicate which ones are more vulnerable. Combining Figure 2.2 and 2.3, the security index can illustrate the vulnerable measurements in a power network.

The values of security index under combined attacks are smaller than the ones under FDI attacks when $C_A < C_I$ from Figure 2.3. For instance, in order to corrupt measurement $j = 10$, the FDI attack needs a value of 11 for attack resources (i.e. a 11-tuple FDI attack) while the combined attack only needs a value of 6 (i.e. a (1,10)-tuple combined attack). This implies that SE is more vulnerable to combined attacks with less attack resources. The results also show that $k_f = 1$ for the combined attacks and the optimal attack cost is $C_I + (\beta_j^\star - 1)C_A$ for the case of $C_A < C_I$, which is consistent with Proposition 2.3.5.

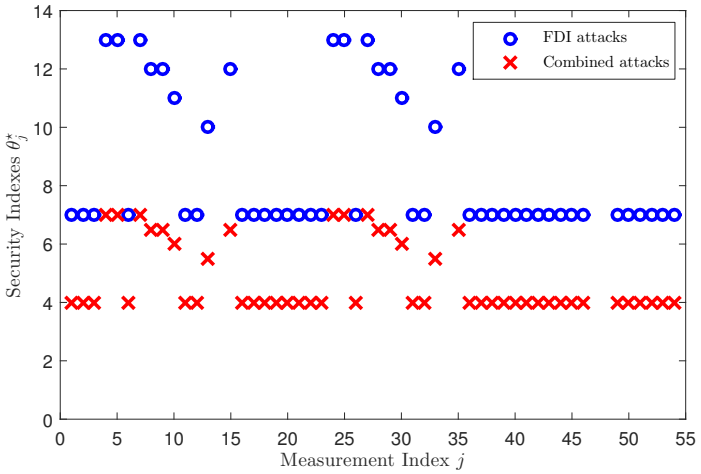


Figure 2.3: The security index $\gamma_j^{a,d}$ under combined attacks and γ_j^a under FDI attacks are plotted versus the measurement index j. Here the cost of FDI attack on per measurement is assumed to be 1 and $C_A = 0.5$ as $C_A/C_I = 0.5$.

Table 2.2: The list of the attacked measurements by integrity or availability in the (1,10)-tuple, (2,9)-tuple and (6,5)-tuple combined attacks.

	Integrity	Availability
(1,10)-tuple attack	10	9, 15, 29, 30, 35, 44, 45, 46, 47, 49
(2,9)-tuple attack	10, 15	9, 29, 30, 35, 44, 45, 46, 47, 49
(6,5)-tuple attack	9, 10, 15, 29, 35, 45	30, 44, 46, 47, 49

2.6.2. DETECTABILITY OF ATTACKS WITH LIMITED KNOWLEDGE

Using the attack policy $f = \tilde{C}_a \Delta X$ for combined attacks and $f = \tilde{C} \Delta X$ for FDI multivariate attacks with the same given model uncertainty, the detection probability of attacks can be obtained according to Section 2.4.2. From Theorem 2.4.2, the compromised measurements set from the optimal solutions of (2.12) w.r.t. the "perturbed" model $\tilde{C}$ in (2.17) is in the same critical tuple with the one w.r.t. C. Thus a set of 11 measurements (a critical tuple) containing measurement $j = 10$ needs to be compromised by the attacker from the security index results in Figure 2.3. This critical tuple includes 11 measurements with index 9, 10, 15, 29, 30, 35, 44, 45, 46, 47, 49, and from Table 2.1 it is known that these measurements are power flows on branches from Bus 4 to 9, Bus 5 to 6, Bus 7 to 9 and power injections on Bus 4, 5, 6, 7, 9. For the sake of comparison, the combined attacks and FDI multivariate attacks are per-

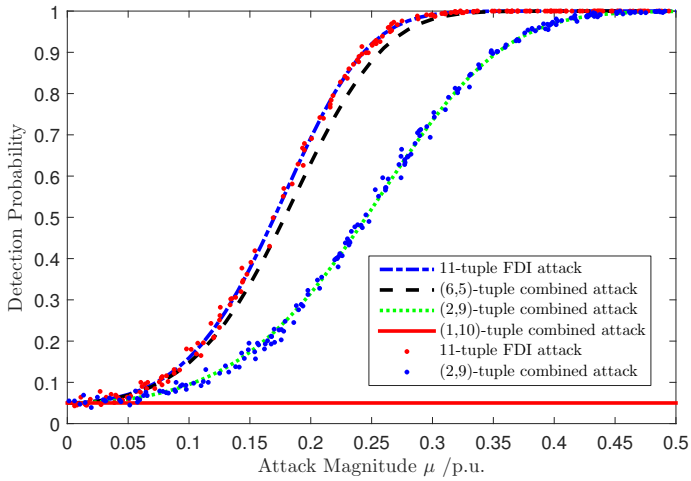


Figure 2.4: The detection probability is plotted versus the attack magnitude. The theoretical results are compared with the empirical detection probability (in red dot for the 11-tuple FDI attack and in blue dot for the (2,9)-tuple combined attack). The attacks are all under structured uncertainty model (error on the model parameters of ±20%) and performed in the same set of 11 measurements (see Table 2.2) and the false alarm rate α is 0.05.

formed in the same set of these 11 measurements. Figure 2.4 shows the detection probability of combined attacks and FDI attacks targeting these 11 measurements, with the structured model uncertainty (error on the line parameters of ±20%). In addition to the theoretical results, the empirical detection probability results are also presented in Figure 2.4 for the 11-tuple FDI attack and (2,9)-tuple combined attack respectively. Table 2.2 lists the measurements attacked by integrity or availability in each combined attack of Figure 2.4. Figure 2.5 shows the detection probability of combined attacks and FDI attacks with different levels of model uncertainty.

To obtain the empirical detection probability in Figure 2.4, Monte Carlo simulations are used. Taking the (2,9)-tuple combined attack as an instance, 200 different values of attack magnitude μ were taken in random from 0 to 0.5 p.u. and the corresponding attack vectors were built. For each attack vector with the taken magnitude μ, total 1000 Monte Carlo runs [2] were executed to obtain the detection probability of such attack. In each Monte Carlo simulation, the measurements were created by the simplified DC power flow calculations with Gaussian noise and corruptions by

[2]The simulations compute the detection probability by counting the proportion of residual signals in each case that exceed the threshold (set in (2.6)). Such simulations require processing of many signals, and the number of 1000 here is set based on empirical observations.

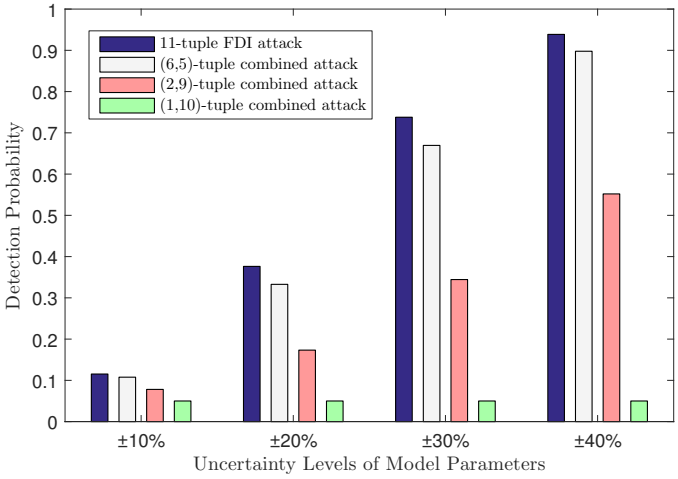


Figure 2.5: The (theoretical) detection probability is plotted versus different levels of model uncertainty (error on the model parameters of ±10%, ±20%, ±30%, ±40%, respectively). The combined attacks and FDI attacks are performed in the same set of 11 measurements (see Table 2.2) and the attack magnitudes are all chosen to $\mu = 0.15$ p.u.. The false alarm rate α is still 0.05.

the attacks. For the attacked measurements, the SE and BDD with the false alarm rate 0.05 were executed. From Figure 2.4 it can be seen that the empirical results of detection probability follow the theoretical ones from (2.15)(2.16). Thus it is feasible to use the theoretical detection probability results for cyber risk analysis in the following. The results in Figure 2.4 illustrate that combined attacks can have lower detection probability comparing with FDI attacks, meaning that SE is more vulnerable to combined attacks as they have higher probability not to be discovered by the BDD. An interesting result is that with smaller k_f the combined attack also has lower probability to be detected. In the case that $k_f = 1$ and $k_a = 10$, the $(1, 10)$-tuple combined attack can keep stealthy, which is consistent with Proposition 2.4.3. The results in Figure 2.5 show that, with different levels of model uncertainty, the detection probability of attacks would increase when the error on the transmission line parameters gets more significant. This can be expected as the attacker has even less knowledge to build attack vectors. Besides, combined attacks still have advantages in keeping stealthy as they can have lower detection probability especially the combined attacks with smaller k_f, and the undetectability of the $(1,10)$-tuple combined attack is independent of parameter uncertainties as discussed in Proposition 2.4.3.

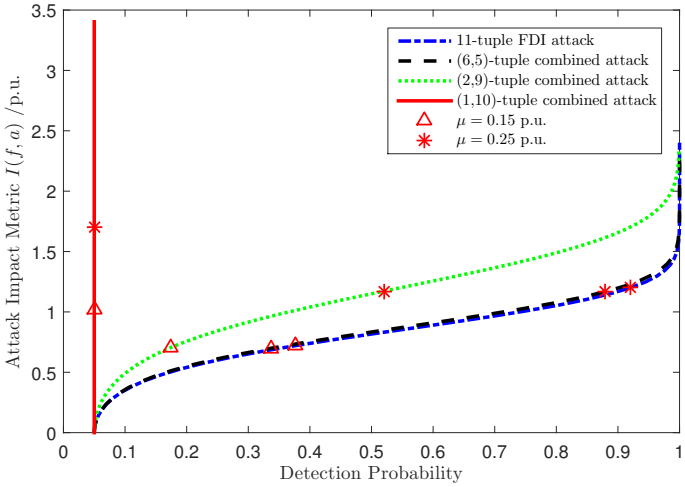


Figure 2.6: The attack impact metric is plotted versus the (theoretical) detection probability. The attacks are all under structured uncertainty model (error on model parameters of ±20%) and performed in the same set of 11 measurements (See Table 2.2). Here it is assumed that $C_A = C_I = 1$ and the false alarm rate α is 0.05.

2.6.3. CYBER RISK METRICS OF DATA ATTACKS

This section continues with the cyber risk analysis of combined and FDI multivariate attacks. Simulations were conducted on the same scenarios as Section 2.6.2 where the attacker manipulates the set of 11 measurements (a critical tuple). The attack impact is analyzed and the risk of the combined attacks and FDI attacks is presented. For the risk analysis, cost values satisfy $C_A = C_I = 1$, thus the security indexes $\tilde{\theta}_j^\star$ w.r.t. $\tilde{C}$ in (2.17) of these attacks are equal to each other and the probability $P(f,a)$ can be "normalized" as discussed in Section IV-B. First, for the attacks with specific model uncertainty (error on the transmission line parameters of ±20%), the results of attack impact metrics versus detection probability are given in Figure 2.6, and the values of risk metrics for combined attacks and FDI attacks versus attack magnitude are shown in Figure 2.7. Second, for the attacks with different levels of model uncertainty, Figure 2.8 also presents the risk metric values of combined attacks and FDI attacks.

Under the perturbed model with uncertainty, the attacker has the possibility to be detected by the BDD while introducing errors on load estimate. From Figure 2.6, it can be concluded that combined attacks can have similar attack impact metrics with FDI attacks but lower detection probability with the same attack magnitude μ

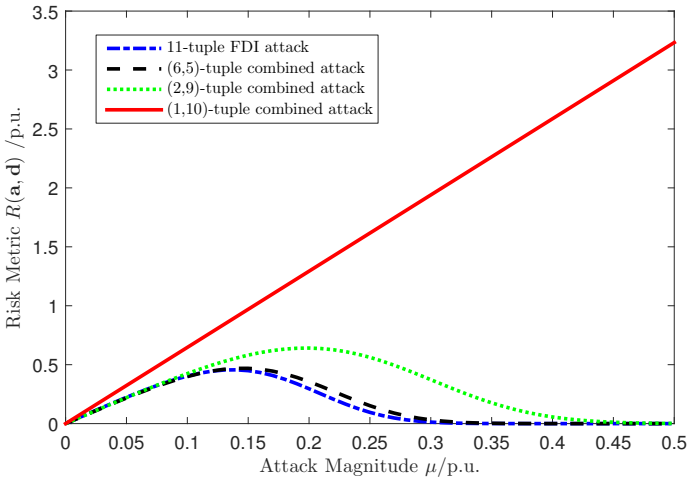


Figure 2.7: The risk metric is plotted versus the attack magnitude. The attacks are all under structured uncertainty model (error on model parameters of ±20%) and performed in the same set of 11 measurements (See Table 2.2). Here it is assumed that $C_A = C_I = 1$ and the false alarm rate α is 0.05.

(0.15 p.u. or 0.25 p.u. as shown in Figure 2.6). Especially the (1,10)-tuple combined attack has larger impact metrics than all other attacks with limited knowledge for the cases that attack magnitude $\mu = 0.15$ p.u. or $\mu = 0.25$ p.u.. For the risk metrics in Figure 2.7, when the attack magnitude μ increases, the risk metric increases due to the low detection probability. After μ reaches certain values, the risk metric decreases since the attacks can be discovered by BDD with high probability. It's also shown that combined attacks can have larger risk metrics especially the cases of (1,10)-tuple and (2,9)-tuple combined attacks. It should be noted that though it is assumed that $C_A = C_I$ to obtain the risk metrics, the risk prioritization of these attacks in Figure 2.7 would not change if $C_A < C_I$ is taken. This is because the combined attacks can be launched with less attack resources when $C_A < C_I$, resulting in larger risk values comparing with FDI attacks. Figure 2.8 illustrates that with bigger errors on the model parameters, the risk metrics would decrease for most of the attacks, meaning that the system faces less risk when the attacker has large model uncertainty as a result of analyzing an out-dated or estimated model in computing attack vectors. From Figure 2.8 we can see, combined attacks with smaller k_f would bring more risk and the (1,10)-tuple combined attack has the largest risk metric under each level of model uncertainty This is due to the fact that this attack can always

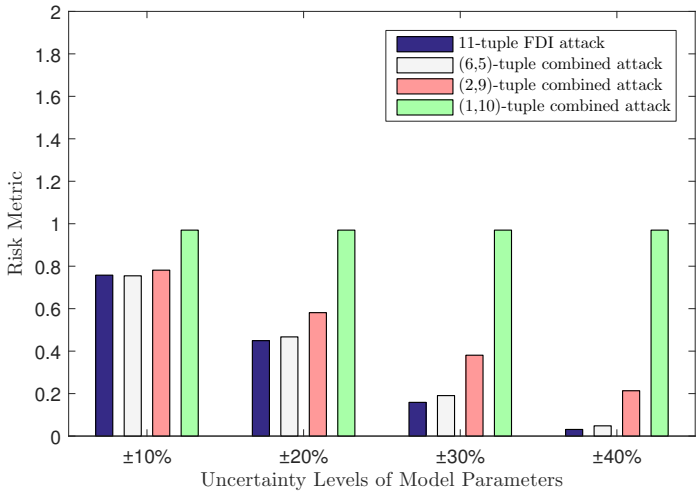


Figure 2.8: The risk metric is plotted versus different levels of model uncertainty (error on the model parameters of ±10%, ±20%, ±30%, ±40%, respectively). The attacks are performed in the same set of 11 measurements (See Table 2.2) and the attack magnitudes are all $\mu = 0.15$ p.u.. It is assumed that $C_A = C_I = 1$ and the false alarm rate α is 0.05.

Table 2.3: Computation time of security index in the IEEE benchmarks

	14-bus	39-bus	118-bus
Time	4.2 s	25.6 s	117 s

keep stealthy even with limited adversarial knowledge.

2.6.4. FURTHER DISCUSSIONS

COMPUTATIONAL EFFICIENCY

In this chapter the big M method is used to express the security index problem as a program of MILP. To show the computation time of this method, security indexes were calculated for IEEE 14-bus, 39-bus and 118-bus systems, all of which have full measurements for the sake of comparison. Note that the big M method does not need the full measurements assumption. The computation time for these four IEEE benchmarks is listed in Table 2.3. The computation was performed on a PC with 3.5 GHz CPU and 8 GB of RAM. The MILP problems were solved using the CPLEX solver for Matlab where the execution time of the algorithm for calculating all the security indexes of each IEEE benchmark was recorded.

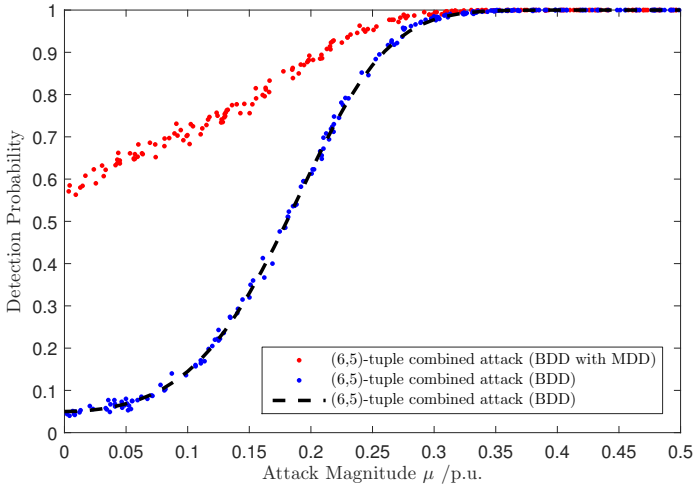


Figure 2.9: The detection probability is plotted versus the attack magnitude. The same (6,5)-tuple combined attack from Figure 2.4 is tested in both cases: one with BDD and MDD, the another one with BDD (without MDD). For MDD test, p_0 is assumed to be 0.06.

Table 2.3 shows that when the system becomes larger, the computation time increases. The MILP formulation imposes challenges for computing security index in large-scale power systems. However, this method could be used off-line for vulnerability assessment. Faster computation time can be achieved on the expense of accuracy using relaxations (such as 1-norm relaxation providing an overestimate of the security index [16]) or some assumptions (such as the full measurements assumption used in the min-cut algorithm [57, 59]).

EXISTENCE OF A DETECTOR FOR AVAILABILITY ATTACK

It should be noted that the previous results of this chapter assume that the SE treats the availability attacks as missing data and no additional alerts are triggered. Although the typical BDD scheme fails to detect availability attacks, a new diagnosis tool could be designed to detect combined attacks. Here an missing data detection (MDD) scheme is proposed. It is assumed that, under normal conditions each measurement may be missing with a given small probability. In particular, let us say that the i-th measurement is missing if $u_{(i)} = 1$, where $u_{(i)} \in \mathbb{B}$ is a Bernoulli distributed random variable with $P(u_{(i)} = 1) = p_i$. The Bernoulli distributed random variables $u_{(i)}$ for $i = 1, \ldots, m$ are assumed to be independent and identically distributed, with

$p_i = p_0$ for all i. The missing data due to abnormal conditions can be detected based on the random variable $u \in \mathbb{B}^m$. Parameterizing $u_{(i)}$ as $u_{(i)} \sim B(p)$, it becomes to test the hypothesis $\mathcal{H}_1$ with a null hypothesis $\mathcal{H}_0$. If $\mathcal{H}_0$ is accepted, that means there is no availability attack and alternatively availability attack exists:

- $\mathcal{H}_0 : p \leq p_0$;
- $\mathcal{H}_1 : p > p_0$.

In other words, it comes to differentiate between cases of low probability of missing data, versus cases where missing data occurs with higher probability. Defining the auxiliary statistic $n_u \triangleq \sum_{i=1}^{m} u_{(i)} = \mathbf{1}^\top u$ which corresponds to the number of missing measurements, we know that n_u follows a binomial distribution, namely $n_u \sim B(m, p)$ with the likelihood function $L(p; u) \triangleq \frac{n_Y!}{(n_Y - n_u)! n_u!}(1-p)^{n_Y - n_u} p^{n_u}$. Thus the statistical test for rejection $\mathcal{H}_0$ is

$$n_u > \bar{\tau}_u,$$

where $\bar{\tau}_u$ is computed to bound the probability of false-alarm of the statistical test.

Recall the current BDD scheme in SE described in Section 2.2.2. If the above MDD scheme is implemented along with the BDD mechanism, the detection probability of combined attacks can be obtained. Note that the random variables $r_{f,a}$ in (2.14) and n_u are not independent since the unavailable measurements will influence the degrees of freedom and the covariance matrix of the residual vector $r_{f,a}$. Thus it's difficult to express the whole detection probability of combined attack under these two detectors mathematically. Monte Carlo simulations can be used instead. For each taken attack magnitude, the given combined attack was implemented through 1000 Monte Carlo runs while in each run the measurements were generated with random errors. If this combined attack triggered any alert on these two detectors, it is said that the attack was detected. Here Figure 2.9 shows the detection probability of (6,5)-tuple combined attack (from Figure 2.4) when the proposed MDD is equipped with the typical BDD. The results show that the MDD could help in detecting the combined attacks.

AC POWER FLOWS

In this chapter for the first time the combined attacks under limited knowledge are explored and cyber risk analysis on the combined or FDI multivariate attacks is con-

ducted. The chapter focuses on establishing the concept of vulnerability and risk of these attacks and exploring this concept in the DC state estimation at the EMS of control. It is expected that this can be a stepping stone towards addressing vulnerability and risk of combined attacks in the AC power flows model.

The combined or FDI attacks explored in this chapter would naturally be more complex to compute under the AC model. In the case of AC state estimation, an attacker would need to have a better knowledge of the system and its operating state. The detection probability of the combined or FDI attacks constructed based on the simplified DC model will be higher and the risk of a successful attack will be lower. Thus, the results of this chapter cannot be directly extrapolated to the case with AC state estimation. However, it is believed that the proposed formulation can be used to explore the AC case by replacing the simplified DC model (matrix C) with a linearization of the AC nonlinear power flow model at a given system state of interest.

3

Co-simulation for Cyber Security Analysis of Data Attacks

To support vulnerability assessment of data attacks, in addition to analytic methods, a platform integrating power system and ICT infrastructure simulators for cyber security tests needs to be developed. This chapter introduces co-simulation techniques to analyze the data attacks on intelligent power grids. First, the analytic approach developed in Chapter 2 is extended to characterize data attacks as optimization programs with the objective specified as security index and constraints corresponding

This chapter is based on the following published work:

[66] K. Pan, A. Teixeira, C. López, & P. Palensky (2017) Co-simulation for Cyber Security Analysis: Data Attacks against Energy Management System. In *8th IEEE International Conference on Smart Grid Communications*, 253–258. DOI: 10.1109/SmartGridComm.2017.8340668;

[67] P. Palensky, A. van der Meer, C. López, A. Joseph, & K. Pan (2017). Cosimulation of Intelligent Power Systems: Fundamentals, Software Architecture, Numerics, and Coupling. *IEEE Industrial Electronics Magazine*, 11(1), 34–50. DOI: 10.1109/MIE.2016.2639825;

[68] P. Palensky, A. van der Meer, C. López, A. Joseph, & K. Pan (2017). Applied Cosimulation of Intelligent Power Systems: Implementing Hybrid Simulators for Complex Power Systems. *IEEE Industrial Electronics Magazine*, 11(2), 6–21. DOI: 10.1109/MIE.2017.2671198.

to the communication network properties. Then the coupling of power system and ICT infrastructure simulators for cyber security analysis is investigated. Finally, the developed tool in the form of co-simulation - coupling the power system simulator DIgSILENT PowerFactory with the communication network simulator OMNeT++, and Matlab for EMS applications (state estimation, optimal power flow, etc.) is presented. Results from the analytic vulnerability assessment are used to conduct attack simulations on the co-simulation platform for a test case. The results indicate that power systems are still more vulnerable to combined attacks and multi-path routing scheme can be used for attack mitigation, and co-simulation can help in validation of the vulnerability and attack impact evaluation.

3.1. INTRODUCTION

The integration of power systems, automated devices, and ICT gives the modern grids the character of a cyber-physical system [69]. As in any cyber-physical system, the power network and its components and the ICT infrastructure are two parts of a larger, heterogeneous system. Combining ICT with the power system may also lead to a number of dependencies that require attention [70]. One important example is security analysis. Typically, contingencies in power systems are considered as independent events, such as the loss of electric components. However, intentional cyber attacks and vulnerabilities from the ICT domain break this assumption as ICT assets could be used to cause damages to several electric components in a coordinated manner [71]; recall the stealthy multivariate attacks against the SE process.

To secure our intelligent power grids, a considerable amount of work has been done on vulnerability assessment of data attacks against energy management systems (EMS), as discussed in Section 2.1.1. In general, these are system-theoretic measures based on analytic methodologies. The work in Chapter 2 indeed provides such an approach for vulnerability assessment. It should be noted that analytic methods may have to ignore some details [1] when modeling the heterogeneous cyber-physical power system, but could be used to characterize attack scenarios and guide the cyber security experiments on testbeds. Tools that integrate cyber and physical components are needed to support the cyber security analysis, from vulnerability assessment to attack impact evaluation and even mitigation schemes development. Besides, in addition to system-theoretic ones, another group of measures come from ICT-specific security. Examples of such measures are firewalls, network intrusion detection systems and authentication, etc. Recently, some organizations such as National Institute of Standards and Technology (NIST) and North American Electric Reliability Corporation (NERC) have proposed security standards (e.g., the recently published IEC 62351-12 [72]) that combine the measures from both ICT-specific and system-theoretic sides [25]. Thus from the above discussion, the following remarks can be concluded:

- The system-theoretic measures based on analytic methods need empirical re-

[1] For instance, the analytic approach developed in Chapter 2 only concerns the number of measurements that an attacker may have to compromise, while the measurement routing topologies are not considered. Of course, it might be the case that in vulnerability analysis not all details can be always included in practice.

sults for validation and further analysis;

- The vulnerability assessment of data attacks on power systems should take attack impact into account, especially the impact evaluated from simulations;
- To improve the security of intelligent power grids, there is a necessity to explore the interactions between system-theoretic and ICT-specific measures.

To support the cyber security analysis above, an integrated platform using various tools including simulators for power systems, SCADA communication networks and EMS applications could offer these capabilities. Co-simulation is currently one of the most popular methods to analyze the behavior of intelligent power grids. Therefore this chapter aims to extend the analytic vulnerability assessment framework in Chapter 2 to incorporate communication network properties and enable them with support from a co-simulation platform.

3.1.1. A REVIEW ON CO-SIMULATION OF INTELLIGENT POWER GRIDS

This section provides a review on the co-simulation of power systems and ICT infrastructure. Noteworthy applications of co-simulation related to intelligent power grids are the analysis of wide area monitoring and control [80], control and optimization in distribution networks [81, 82], and distributed energy integration [83–85]. In such applications, co-simulation can conveniently scrutinize interactions between systems of completely different natures. For instance, the impact of communication latency on the power system has been analyzed in [86], while the impact of cyber attacks has been studied in [23, 24]. Co-simulation has also proven to be useful to explore artificial intelligence (AI) applications in power systems [87]. Real-time/hardware-in-the-loop (HIL) testbeds have been proposed in the work [88] for automation-related co-simulations. Setups like in [22] are currently used for evaluating the impact of latency or packet loss on smart grid control applications.

Over the past decade, profound efforts have been made to couple continuous power system simulators with discrete communication network simulators. The electric power and communication synchronizing simulator (EPOCHS) [73] is one of the first, and it combines power system simulators with instances of network simulator 2 (ns-2) at run time. The global event-driven co-simulation framework (GECO) [76] for evaluation of wide area monitoring and control methods integrates

Name	Application	Components	Synchronization
EPOCHS [73]	Protection and control	PSCAD, PSLF, and ns-2	Point-based
OpenDSS & OMNet++ [74]	Wide area monitoring and control	OpenDSS, OMNet++	Point-based
Adevs+ns-2 [75]	Wide area monitoring and control	Adevs, ns-2	Event-driven
GECO [76]	Wide area protection and control	PSLF, ns-2	Event-driven
Greenbench [23]	Cyber security in distribution grid	PSCAD, OMNet++	Event-driven
PowerNet [77]	Monitoring power grid devices	Modelica, ns-2	Master-slave
VPNET [78]	Networked power converter system	VTB, OPNET	Master-slave
INSPIRE [69]	Monitoring and control	DIgSILENT PowerFactory, OPNET	Master-slave
OpenDSS & ns-2 [79]	Distributed energy resources integration	OpenDSS, ns-2	Not addressed
TASSCS [24]	Cyber security of SCADA system	PowerWorld, OPNET	N/A

Table 3.1: Examples of co-simulation of power systems and ICT infrastructure. N/A: not available. OpenDSS: Open Distribution System Simulator; OMNeT++: Objective Modular Network Testbed in C++; TASSCS: testbed for analyzing security of SCADA control system; PSCAD: power system computer-aided design; VTB: virtual test bed; OPNET: optimized network engineering tools; DIgSILENT: digital simulation and electrical network.

PSLF with ns-2. GECO runs globally in a discrete event-driven manner whereas a global event scheduler is used to handle power system iteration events and communication network events. The integrated co-simulation of power and ICT systems for real-time evaluation (INSPIRE) [69] uses the High Level Architecture (HLA, IEEE 1516) for time management, providing a co-simulation platform for modeling the effects of ICT infrastructures on power systems. Table 3.1 provides a non-exhaustive

list of examples of co-simulation of power systems and ICT infrastructure. Time synchronization mechanism in Table 3.1 would be further discussed in Section 3.3.2.

Co-simulation testbeds have also been developed to design and evaluate cyber security aspects particularly, e.g., Nationnal SCADA TestBed [89], Virtual Control System Environment [90], Virtual Power System Testbed [91], and the Testbed for Analyzing Security of SCADA Control Systems (TASSCS)[24]. In [92], a comprehensive testbed for modeling and simulating practical cyber-events is developed. It uses Real Time Digital Simulator (RTDS) and ns-3 for power system and communication network simulations respectively. Real Phasor Measurement Units (PMUs) and Phasor Data Concentrator (PDC) are integrated into the testbed to collect phasor measurements. The cyber-defense experimental tool DeterLab can launch, monitor and analyze security events. In [4], the offline and real-time testbeds are proposed for power system substations cyber security. The offline testbed uses open source software packages. The core of this testbed is the co-simulation of OpenDSS (Open Distribution System Simulator) and MATLAB. Using the Component Object Model (COM) interface, MATLAB can access to the results and parameters of OpenDSS. The cyber intrusions like False Data Injection (FDI), Denial of Service (DoS), are implemented in MATLAB. The real-time security testbed is based on DIgSILENT PowerFactory for power system simulation. The components in each substation are mapped with the Object Linking and Embedding for Process Control (OPC) client and linked to an OPC server. Intrusion detection system and firewalls are also deployed in substations. This testbed can be used to explore how a real cyber attack works. From the literature, most of the testbeds are used for impact analysis of cyber attacks. To be noted, co-simulation can also be used for other research aspects, such as vulnerability assessment, mitigation measures development, security validation and interoperability.

3.1.2. CONTRIBUTIONS AND OUTLINE

The two parts of work in analytic assessment and numerical simulations for cyber security analysis are usually conducted independently even though they are related. This chapter aims to contribute in closing the gap by extending the analytic vulnerability assessment framework in Chapter 2 to incorporate communication network properties and developing a co-simulation platform to conduct simulations on data

attacks against EMS. The contributions are as follows:

(i) The proposed vulnerability assessment method in Section 2.3 is extended to incorporate communication network properties, for a better characterization of cyber-physical power systems. Routing vectors and matrices are proposed to model the power system communication network with a particular focus on the topology and data routing schemes. Finally the security index quantifying the minimum attack resources needed by the attacker on communication nodes and links is presented through (3.3) and Theorem 3.2.1 for computation solutions. It is shown that power systems are still more vulnerable to combined attacks and multi-path routing can be adopted for attack mitigation.

(ii) After a thorough discussion on coupling power system and ICT infrastructure, including simulators integration and time synchronization methods, a real-time co-simulation platform is developed for the cyber security analysis of EMS in power systems. It is accomplished with the integration of simulators such as DIgSILENT PowerFactory for power system, OMNeT++ for communication network and Matlab for EMS. A customized scheduler is implemented as a master algorithm to coordinate the simulators and run in real-time. We use the attack scenario from the analytic approach to conduct simulations on the developed platform and analyze the simulation results.

Section 3.2 proposes routing vector and matrix for communication network modeling. The analytic vulnerability assessment is then extended to incorporate the properties of communication topology and routing schemes. Section 3.3 investigates the coupling of power system and ICT infrastructure. The modeling challenges, time synchronization of continuous and discrete simulators, and real-time co-simulation are illustrated. In Section 3.4, the co-simulation platform is presented in detail, including how the power system and communication network are modeled, how the tools are integrated and how the attacks are implemented in OMNeT++. Section 3.5 first shows the results of security index of combined attacks under communication model with single or multi-path routing scheme. Then the results from co-simulations are also presented and analyzed. To the end, a brief discussion is provided on combining system-theoretic and ICT-specific measures to protect power systems from cyber attacks.

3.2. VULNERABILITY ANALYSIS INCORPORATING COMMUNICATION PROPERTIES

SCADA systems are vulnerable to a large number of cyber threats. As shown in Figure 2.1, the manipulation of measurements can arise from various levels (A1: RTU sensors, A2: SCADA communication network, A3: SCADA control center). To conclude from the analysis in Chapter 2, the measurements $Y \in \mathbb{R}^{n_Y}$ under different data attacks from the view of State Estimation (SE) can be presented as follows:

- Data integrity attack - known as false data injection (FDI) attack, is able to change measurements from Y to $Y + f$ where $f \in \mathbb{R}^{n_Y}$ is the FDI attack vector;
- Data availability attack - includes DoS or jamming attack which would make specific measurements unavailable to SE, i.e., $Y_a = (I_{n^Y} - \text{diag}(a))\,Y$ where $a \in \{0, 1\}^{n_Y}$ is the availability attack vector and I is an identity matrix;
- Combined attack - launches the FDI attack and availability attack simultaneously that makes the measurements from Y to $(I_{n^Y} - \text{diag}(a))\,Y + f$.

As discussed in the preceding Section 2.2 and Section 2.3, vulnerability assessment of data attacks is presented through the notion of *security index* which concerns the level of efforts required by the attacker to keep stealthy from the bad data detection (BDD) mechanism. It has been shown that if the attacker corrupts certain measurements using FDI attack vector $f = C\Delta X$, where C represents the network model in (2.1), it can remain stealthy but perturb the current state to a degree of ΔX. It's also shown in our work [36] that combined attacks can achieve the same target with the attack vector $f = (I_{n^Y} - \text{diag}(a))C\Delta X$. In sight of this, the security index problem of (2.10) was considered that the objective $\beta_j^\star := \min_{f,a} \|f\|_0 + \|a\|_0$ illustrates how many measurements to be manipulated and the constraints correspond to stealthiness conditions and capability limits for combined attacks.

3.2.1. COMMUNICATION ROUTING SCHEME MODELING

The vulnerability assessment methodology in Chapter 2 suits for the case that attacks arise from the level of A1 in Figure 2.1. This security index directly demonstrates that manipulations on several RTU sensors are required by the attacker. However, it does not consider the model of power system communication network and

hence lacks of details for modeling a cyber-physical power system. Besides, in practice, tampering RTUs directly becomes much harder as more of them are authenticated and secured. A more interesting scenario is to look into attacks from the level of A2 since usually the attacker would exploit vulnerabilities more in communication networks, e.g., compromising remote access points, gaining access to corporate networks. However, modeling the communication network in an analytic framework is challenging due to its complexity and heterogeneity. Here, the communication network properties of interest for security analysis are:

- Communication network topology;
- Routing schemes - the routing paths of data packets;
- Communication latency - the delay that happens in data communication;
- Packet loss or missing data - the possibility of packet loss in communications.

In what follows an approach is proposed to tackle with the first two properties that can be characterized in the security index problem formulations. Another two properties of communication networks, latency and packet loss, could also be incorporated into analytic framework, not for vulnerability assessment but for combining ICT-specific and system-theoretic measures. First, let us recall the SCADA communication network in power grids (Figure 1.1). Wide area networks (WANs) are deployed to deliver measurements from substations to the control center. The communication lines are usually laid along with the transmission lines between substations with the cables installations. Thus, the measurements sent from a substation would go through several substations, where switches, routers and multiplexers multiplex the data from different substations onto the communication link [47].

With the knowledge above, it is feasible to represent each substation as a communication node that receives and transmits data. With communication links between nodes, mesh topology is used to improve utilization of available infrastructures. This network is a multi-hop network where data packets would be routed through multiple nodes before reaching the SCADA control center [93]. Figure 3.1 shows the communication model of IEEE 14-bus system. There are 10 nodes and 15 communication links on the 14-bus system's WAN. Each node and link represents one substation and one communication line in a physical system. Here it is assumed that the control center is located at the reference bus, i.e., bus 1/node 0.

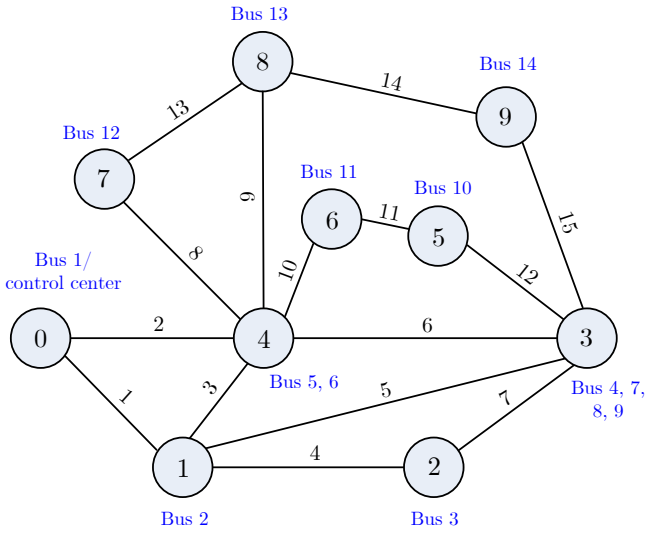


Figure 3.1: Communication model of the IEEE 14-bus test system based on [93].

After accessing one node, the attacker can implement data integrity attacks on some measurements that are collected on this node or routed through it, by compromising the substation network or sensors. The adversary may also use various data availability attacks on these measurements, by jamming the substation network, launching DoS attacks on the substation server, router, switches or multiplexer [47]. However, if the attacker manipulates a communication link, he may only launch availability attacks on the measurements that traverse this link, instead of corrupting the data integrity.

Next in the following, a concept of routing vector and matrix is introduced for the communication network modeling. We can describe the model in Figure 3.1 as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of connected communication links. For each node $n \in \mathcal{V}$, single or multi-path routing schemes can be implemented. Hence, any measurement j can have single or multiple routes to the control center. A binary vector called *routing vector* is defined for each route of measurement j,

$$r_{j,p} := \left[r_{vj,p}^{\top} \quad r_{ej,p}^{\top} \right]^{\top}, \tag{3.1}$$

where $r_{j,p}$ denotes the routing vector for the p-th route of measurement j. $r_{vj,p} \in \{0,1\}^{|\mathcal{V}|}$ represents the vector corresponding to nodes in this route and the entries

are 1 if this route traverses the corresponding nodes. Here $|\mathcal{V}|$ denotes the cardinality of the set $\mathcal{V}$. Similarly, $r_{ej,p} \in \{0,1\}^{|\mathcal{E}|}$ is the vector of communication links and the entries are 1 if this route traverses the corresponding links.

For a given communication model, using the graph $\mathcal{G}$ and the routing schemes, all of the routing vectors can be obtained. Then for a number of n_Y measurements in the power system, a binary matrix called *routing matrix* can be established,

$$R_p = \begin{bmatrix} R_v & R_e \end{bmatrix}. \tag{3.2}$$

In (3.2), $R_p \in \{0,1\}^{n_P \times (|\mathcal{V}|+|\mathcal{E}|)}$ denotes the routing matrix for a communication model that has a total of n_P routes, and R_v denotes the matrix corresponding to nodes, i.e., R_v is composed of $r_{vj,p}^\top$ for its rows. Correspondingly R_e represents the matrix of communication links that each row of R_e is the $r_{ej,p}^\top$ vector. Using the routing matrix, the routes of all the measurements are mapped to nodes and communication links. Thus it can be seen that, the proposed *routing vector* and *routing matrix* contain the information of communication topology and routing schemes.

3.2.2. SECURITY INDEX UNDER THE COMMUNICATION MODEL

Similarly, the vulnerability of each measurement to combined attacks can be quantified by computing the minimum number of communication nodes and links that an attacker needs to compromise. Here two binary vectors $v \in \{0,1\}^{|\mathcal{V}|}$ and $e \in \{0,1\}^{|\mathcal{E}|}$ are introduced. If $v(i)$ is 1 then the i-th node is attacked; otherwise $v(i)$ is 0. If $e(i)$ is 1 then a certain link is attacked; otherwise $e(i)$ is 0. It is assumed that integrity attacks can only be made on nodes, while availability attacks can be launched on both nodes and links. Then a security index quantifying the total number of compromised nodes and links by a combined attack can be formulated as

$$\begin{aligned} \delta_j^\star := \min_{\Delta X, a, v, e} \quad & \|v\|_0 + \|e\|_0 \\ \text{s.t.} \quad & f = C_a \Delta X, \\ & C_a = (I_{n^Y} - \operatorname{diag}(a))C, \\ & f(j) = \mu, \\ & f(l) = 0,\ \forall l \in \mathcal{L}, \\ & f(i) = 0 \ \text{ if } r_{vi,p} = 0,\ \forall i \neq j,\ p, \\ & a(k) \leq r_{vk,p} v + r_{ek,p} e,\ \forall k \neq j,\ p, \\ & a,\ v,\ e \text{ are all binary vectors,} \end{aligned} \tag{3.3}$$

where the constraints using the routing vectors map the combined attacks (with f and a) on measurements of the SE process to attacks on communication nodes and links. These two constraints indicate that in order to launch data integrity attack on measurement i, all of its routes should include at least one attacked node. When the attacker launches data availability attack on measurement k, all of its routes should include at least one attacked node or one attacked communication link. To solve this NP-hard problem, the computation solution using the big M approach can still be used for the reformulation of (3.3),

Theorem 3.2.1. *For any measurement index $j \in \{1, \ldots, n_Y\}$ and non-zero attack magnitude μ, the optimization (3.3) can be equivalently described via the following Mixed Integer Linear Program (MILP),*

$$\begin{aligned} \delta_j^\star = \min_{\Delta X, w, a, v, e} \quad & \sum_{i=1}^{|\mathcal{V}|} v(i) + \sum_{k=1}^{|\mathcal{E}|} e(k) \\ s.t. \quad & C\Delta X \le M(w+a), \\ & -C\Delta X \le M(w+a), \\ & C(j,:)\Delta X = \mu, \\ & C(l,:)\Delta X = 0, \ \forall l \in \mathcal{L}, \\ & Aw \le R_v v, \\ & Aa \le R_v v + R_e e, \\ & w, a, v, e \text{ are all binary vectors,} \end{aligned} \tag{3.4}$$

where $A \in \{0,1\}^{n_p \times n_Y}$ is a constant binary matrix mapping all the measurements to corresponding routes. For instance, if single-path routing scheme is implemented, A is an identity matrix.

Proof. This proof is similar to the proof of Theorem 2.3.3 for describing (2.10) via (2.11). It still follows by re-writing (3.3) as (3.4). In addition to the proof in Theorem 2.3.3, it is notable that the constraints using the routing vectors in (3.3) can be equivalently re-written as the constraints using the routing matrices in (3.4). The proof concludes by noting that the objective functions of both programs satisfy the equality $\|v\|_0 + \|e\|_0 = \sum v(i) + \sum e(k)$. ■

Now the optimization programs of security index derived so far can identify the compromised sets of communication nodes and links for combined attacks. Similar to (2.12), the costs of manipulating communication nodes and links can be con-

sidered, while the attack resources needed for compromising nodes and links are usually different. Say that attacking each communication node has the same cost C_V, and the cost of attacking per communication link is C_E. Thus a security index problem can be formulated to quantify the minimum attack costs under the communication model by rewriting the objective of (3.4) as

$$\eta_j^\star := \min_{\Delta X, w, a, v, e} \left\{ \sum_i^{|\mathcal{V}|} C_V v(i) + \sum_k^{|\mathcal{E}|} C_E e(k) \right\}. \tag{3.5}$$

Under the communication models, attacking nodes means gaining access to substations and launching attacks on substation local networks, which implies that the adversary needs to compromise a large number of components. For attacks on communication links, the data flow through the attacked link is disrupted, which can be accomplished by gaining access to substations, jamming or flooding the channels, or even physical disruptions. Thus it is reasonable to assume that attacking nodes costs more than attacking communication links.

The security indexes $\delta_j^\star$ and $\eta_j^\star$ can illustrate the vulnerability of power systems to combined attacks on the communication network. It should be noted that some ICT-specific security measures can be modeled in (3.3). For instance, multi-path routing schemes can be described using routing vectors in constraints of (3.3). Data authentication can be implemented by adding constraints to indicate which measurement originating from the node with authentication is protected.

These analytic vulnerability assessment programs in this section, however, do not consider the attack impact on the physical system operations. In fact, data attacks with the same security index could have considerable different physical impact. Co-simulation could offer the capabilities to look into the attack impact and provide empirical results to validate and contribute in developing mitigation measures, which will be detailed in the following.

3.3. COUPLING POWER SYSTEM AND ICT SIMULATORS

The power network and ICT infrastructure in any power systems are two parts of a heterogeneous system. The use of co-simulation to investigate the mutual influences of ICT and power systems and, therefore, the behavior of intelligent power grids has become significant.

3.3.1. MODELING AND SIMULATION CHALLENGES

A *full simulation* to represent ICT in a simulation setup is considered in this thesis. All ICT elements (e.g., switches) are simulated or imitated with proxy code that uses stochastic or other simplified means of representing the time-domain behavior of the system. Thus, as with all digital systems, communication networks are modeled as a sequence of discrete events (e.g., sending and receiving packets, packet buffer overflows, etc.), while power systems are typically modeled as continuous time functions using differential-algebraic equations (DAEs), although discrete power system events occur as well when the status of breakers, switches, and relays change. Consequently, a holistic model of a smart grid must include both continuous and discrete aspects.

According to [94], simulation paradigms can be divided into three time management categories:

- fixed time step-size simulation in which the simulation time is discretized in equal time steps;
- continuous simulation, which commonly applies adaptive time step-size control approach;
- discrete-event simulation, which advances the simulation time only when the discrete events occur.

Intelligent power grids often need multiple models, which need to fit into heterogeneous simulation paradigms. The ICT part of such a multi-domain system is normally implemented as a discrete-event simulation, while the power system part is included as a continuous or fixed time step-size simulation. As mentioned previously, hybrid simulations can be a solution for this problem, i.e., single solvers that address multiple models [95–97]. However, such methods scale badly and can hence only be used for simple use cases, not for fully-fledged system studies. As touched upon before, co-simulation of intelligent power grids, i.e., hybrid continuous and discrete models with multiple solvers, comes with advantages but also challenges:

- The integration of continuous power system and discrete-event communication network simulations needs sophisticated synchronization mechanisms. The next subsection will present methods to tackle this challenge.

- Error estimation and validation of co-simulation is a challenge. The interdependency of hybrid models from power system and ICT parts makes it hard to identify where the simulation error comes from. Different synchronization methods in co-simulation also impact the simulation accuracy.
- Interoperability of the various simulators requires standardized interfaces (e.g., the High Level Architecture (HLA) and the Functional Mock-up Interface (FMI)).

3.3.2. SYNCHRONIZATION OF DISCRETE AND CONTINUOUS SIMULATORS

When building a co-simulation platform for intelligent power grids, the synchronization mechanism between the subsystems under consideration is one of the performance dominating factors. It has a direct impact on the convergence and accuracy of the simulation results.

Time synchronization between continuous and discrete simulations can happen either conservatively or optimistically. Conservative synchronization guarantees strict processing of logical time. The optimistic alternative allows a violation of the step-by-step processing, but needs additional control mechanisms that could detect and recover violations [98]. The simulators must be capable of rolling back the overall simulation time. Unfortunately, many power system simulators do not possess this functionality [69]. Synchronization methods are mainly subdivided into three categories: point-based, event-driven and master-slave [99].

POINT-BASED

While the simulation of power system dynamics uses a time stepped approach, the communication networks are typically modeled as discrete event systems. One intuitive synchronization method is to use predefined synchronization points. As shown in Figure 3.2, individual simulators run in parallel and stop at the synchronization points to exchange information. The synchronization points are predetermined. However, in most cases, the communication need between two simulators is created by events generated by one of the models, which, in case of ICT models, may even have a stochastic nature [77].

The point-based synchronization method may introduce inaccuracies in the co-simulation. When system output variables need to be exchanged between two synchronization points, both subsystems have to wait until the next synchroniza-

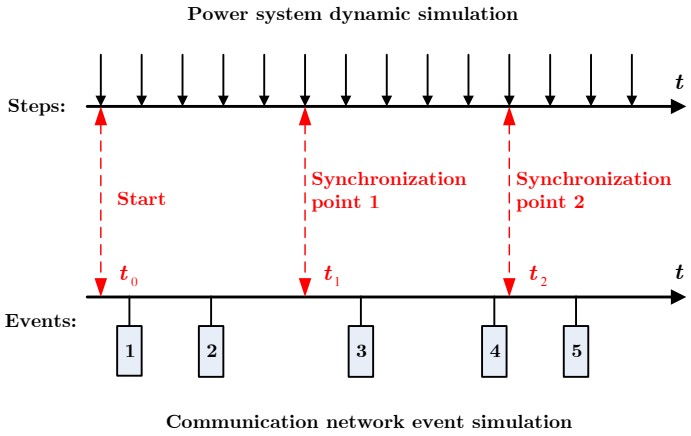


Figure 3.2: Point-based synchronization method.

tion point. This delay introduces error accumulation into the simulation, and possibly impairs the accuracy of the overall simulation results. A simple solution is to reduce the time interval between synchronization points, e.g., exchange data in each time step of power system dynamic simulation [100]. In [74], an advanced point-based approach is proposed, in which the next synchronization point is not predefined but given as a parameter to the continuous power system simulator.

EVENT-DRIVEN

In [76], a global event-driven co-simulation framework is proposed. The event-driven synchronization is shown in Figure 3.3. It treats each iteration round of the continuous power system simulation as discrete events and mixes them with communication network events. All the discrete events form an event queue in chronological order. A global event scheduler checks the event queue and individually handles corresponding control for power system or communication network events. Both simulators can suspend themselves and yield the control back to the scheduler when subsequent events occur. The discrete event specification formalism could be used to model both the power system and the communication network simulation. It provides a rigorous mathematical basis for simulating hybrid system models [75] and is widely used for event-driven synchronization.

Using the event-driven method, the time step-size of the power system simulation significantly impacts the overall co-simulation time. Besides, the interface

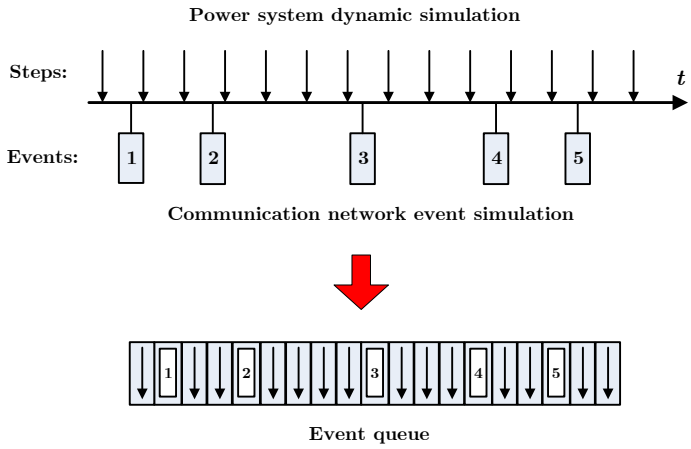


Figure 3.3: Event-driven synchronization method based on [76].

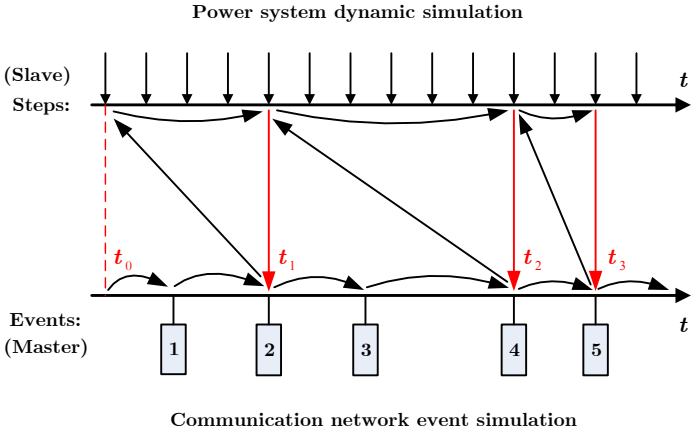


Figure 3.4: Master-slave synchronization method: one simulator acts as the master.

between simulators can be a performance bottleneck, grinding down scalability. In [76], as the system scale grows, the simulation time also increases because of the increased number of interactions in the interface. Hence, the performance is highly dependent on the capabilities of the respective interfaces.

MASTER-SLAVE

The third type of synchronization mechanism, shown in Figure 3.4, is a typical master-slave configuration that allows one simulator (often the discrete-event simulator) as a master simulator to coordinate the entire co-simulation. In Figure 3.4, the com-

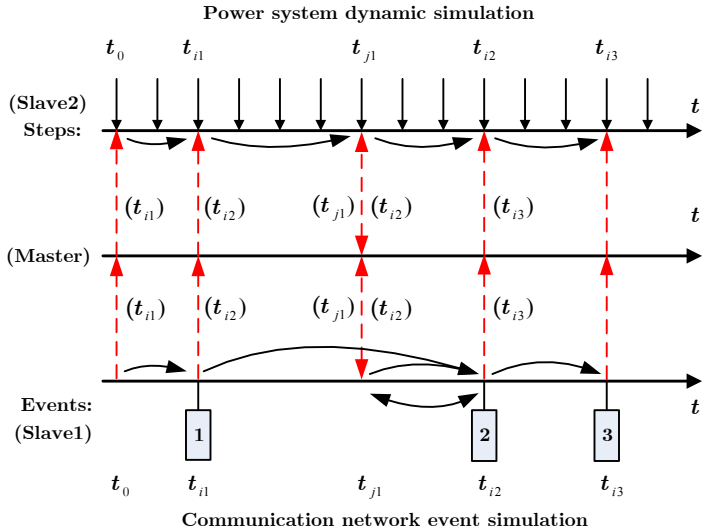


Figure 3.5: Master-slave synchronization method: using a dedicated master component.

munication network simulator (as the master) controls the power system simulator (as the slave) throughout the simulation process. The master starts the simulation at t_0. When the event at t_1 needs the information from the slave, the master coordinates the slave to simulate from t_0 to t_1 and sends data to the master. For the master-slave approach, the synchronization performance is limited by the capabilities of the master simulator. As discussed in [101], the drawbacks are (i) events, generated in the slave cannot be communicated to the master immediately; (ii) execution is typically sequential; (iii) scalability issues inhibit the integration of an arbitrary number of simulators. The latter can be potentially be overcome if one dedicated master algorithm is used to orchestrate all simulators that act as slaves. All slaves have to tell the master when their next event is anticipated. The master then picks the time of the earliest event in this list and declares it to all slaves as the next synchronization point.

Figure 3.5 shows an example where the master tells slave 2 (power system simulator) the time t_{i1}, which is the time of the first event in slave 1 (the communication network simulator). Each slave executes simulation to t_{i1}, where finally data exchange takes place (not shown in the figure for simplicity). The next events are at t_{i2} and t_{j1}. The latter wins since it is earlier, so the next synchronization point is at t_{j1}. Slave 1 has to roll back the simulation time from t_{i2} to t_{j1}. since it normally

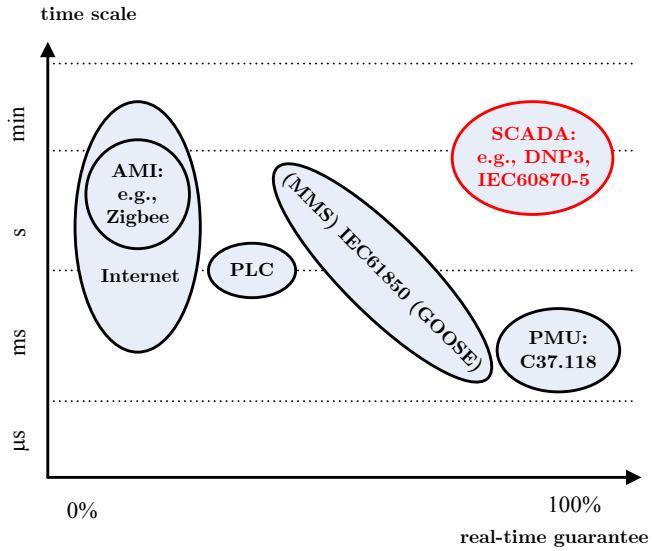


Figure 3.6: Real time guarantees and time scales of power system communication protocols. AMI: Advanced Metering Infrastructure; PLC: power-line communication; MMS: manufacturing message specification; GOOSE: generic object-oriented substation event.

jumps from one event to the next. The scalability of this approach, specially when handling asynchronous events, is a potential challenge on the road.

PRACTICAL CONSIDERATIONS

The majority of recent work on co-simulation platforms focuses on the integration of one power system simulator with one communication network simulator [102]. However, intelligent multi energy systems (e.g., power-to-heat settings with market-integration) need multi-physics capabilities and large scalability. The corresponding co-simulation needs to couple more than one physical systems. For this case, simple synchronization mechanisms work for coupling two simulators but would fail when coupling more. The second master-slave method with dedicated master algorithm shown in this section could be used.

From the preceding, it can be deducted that in many cases the designers have to make a trade-off between accuracy, efficiency, and scalability. It should be noted that the accuracy and efficiency in coupling depends on the simulation tools selected, the interfaces, and the synchronization mechanisms. This also implies the level of control that the designer could have on the simulation tools.

3.3.3. REAL-TIME CO-SIMULATION

ICT, and especially the controls of intelligent power grids, expose another important aspect that the simulation models have to consider: real-time guarantees. As shown in Figure 3.6, some power system protocols offer real-time guarantees, while others operate on best-effort basis (i.e., no communication speed or fidelity is guaranteed). The associated applications either rely on guaranteed latency and throughput or have a more relaxed use. For phenomena that involves loose real-time guarantees and long time scales, it would be possible to choose the interfacing and synchronization methods introduced above. However, for phenomena that involves strict real-time guarantees and short time scales, e.g., PMU based monitoring, it is challenging to design capable interfaces and appropriate synchronization methods. Real time co-simulation based on powerful real-time simulators are needed under such circumstances.

- *RTDS based:* In [70], RTDS is used for power system simulation and ns-3 is used to simulate the communication network. This testbed is mainly for wide area monitoring protection and control research. A testbed is developed in [103] for analyzing the impact of cyber events on microgrids using RTDS as power system simulator and common open research emulator (CORE) as communication network emulator.

- *OPAL-RT based:* OPAL-RT is another platform that supports real-time cosimulation. The Orchestra API acts as the co-simulation scheduler and coordinates the components connected to OPAL-RT. In [104], this real-time co-simulation environment makes use of the compatibility of OPAL-RT and Simulink to develop PMU applications.

- *PowerFactory based:* DIgSILENT PowerFactory, a versatile power system simulator for workstations, also provides a real-time mode. PowerFactory can be interfaced with other hardware or software components through the OPC communication protocol and various application programming interfaces. In [105], the PowerCyber testbed is built using the integration of PowerFactory with RTUs in order to perform cyber-physical security testing.

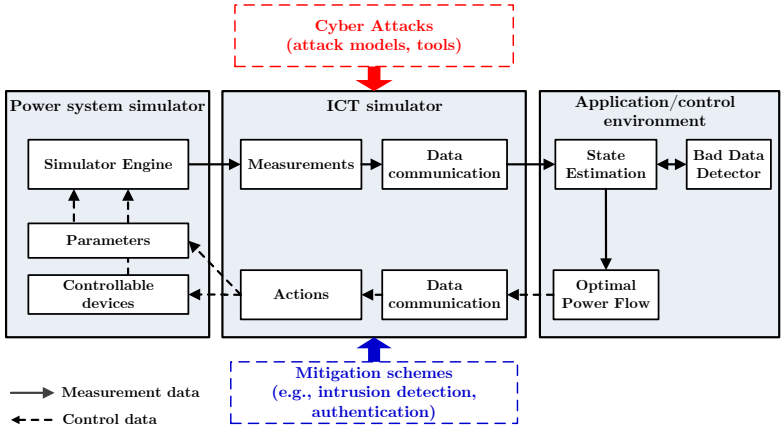


Figure 3.7: Co-simulation based cyber security analysis framework. This figure is adapted from [4].

3.4. CO-SIMULATION FOR POWER SYSTEM CYBER SECURITY ANALYSIS

3.4.1. CO-SIMULATION FRAMEWORK AND TOOLS

A co-simulation framework is an integrated environment including simulators for power system, communication network and application or control. Under the co-simulation, the SCADA communication network can be modeled as a hierarchical one that is close to reality, instead of an abstractive instance in Figure 3.1. Using co-simulation, the attack scenarios from the analytic vulnerability assessment optimization programs in Section 3.2 can be validated and the attack impact can be evaluated from simulation results.

The co-simulation framework for cyber security analysis is shown in Figure 3.7 and is implemented on top of the integration of simulators for power system, communication network and application or control. This platform should have the capabilities of (i) being modular, extensible and flexible to simulate communication networks; (ii) being easy to implement attack scenarios and mitigation measures.

Next the developed co-simulation platform is shown in Figure 3.8. In order to allow for real-time analysis of cyber attacks, the co-simulation is implemented with three simulators: DIgSILENT PowerFactory for the power system, OMNeT++ (Objective Modular Network Testbed in C++) for the communication network, and Matlab/Matpower [106] for the EMS applications. Here, the power flow measure-

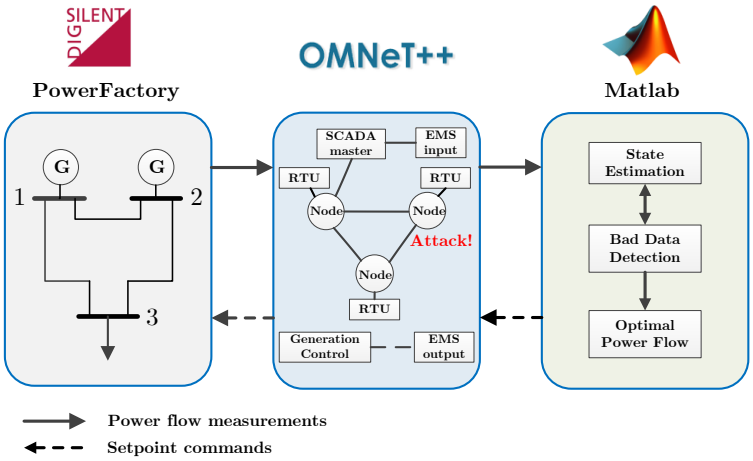


Figure 3.8: Co-simulation diagram with simulators DIgSILENT PowerFactory for the power system, OMNeT++ for the communication network, and Matlab/Matpower for the EMS algorithms.

ments going in and out of each bus of the power system simulated in PowerFactory are sent to the EMS applications in Matlab through a communication network simulated in OMNeT++. OMNet++ is selected because it is a generic simulation engine and it allows plug-n-play through NED (Network Description) language and integration to external devices. Besides, there are various open-source model libraries (e.g., INET Framework for the Internet stack and wired or wireless link layer protocols [107]) that support the communication network modeling in OMNET++.

POWER SYSTEM SIMULATOR

DIgSILENT PowerFactory is used to conduct a consecutive power flow simulation. PowerFactory's Python API is adopted to create a script that controls the execution of the simulation. The same script implements the interface with OMNeT++. Real-time execution is achieved by synchronizing the power flow simulations with the system clock. The script sends measurements to OMNeT++ every fixed time (set to be 5 seconds), but it can expect generator setpoints at any time. Thus, a dedicated thread that received setpoints and sets them in the power system model in PowerFactory is required. This thread sets the generators according to the setpoints when they arrive, unless a power flow calculation is being executed, in which case it waits for the calculation to finish.

COMMUNICATION NETWORK SIMULATOR

OMNeT++ is used for discrete-event based communication network simulation. The communication model of SCADA in OMNeT++ is shown in Figure 3.9. A customized OMNeT++ *scheduler* is built to enable data exchange with DIgSILENT PowerFactory and Matlab over TCP/IP [2] sockets and run OMNeT++ in real-time. With the help of INET Framework for TCP/IP protocols, modules for each SCADA communication components have been developed. In Figure 3.9, the "RTU" is a module served by the scheduler and acts as a RTU proxy. The second module developed called "MTU" works as the master terminal unit and data concentrator that receives data packets and has a FIFO (first in, first out) queue. There is a "Modem" module that acts as a communication bridge, and a *Router* module with a routing table [3] for all the packets. Thus, the RTU, Modem and Router represent the LAN (local area network) of a substation (recall Figure 1.1). Besides, the modules "EMSInput" and "EMSInout" provide measurements to EMS and receive setpoints for generators from EMS applications in Matlab respectively. For the message implementation, a new packet class "MeasurePacket" is derived to contain the measurement data and be used by all the modules and scheduler. Thus, there are two kinds of communication channels: (i) channel of the LAN and (ii) channel of the WAN (wide area network) between routers. Different latency and packet loss probability parameters are set in these two channels. It should be noted that implementation of a real SCADA system with protocols (e.g., DNP3.0, IEC 61850) and hierarchical network structure in OMNeT++ is not the main focus. Instead this thesis tries to explore how co-simulation can support the cyber security analysis.

EMS APPLICATIONS ENVIRONMENT

Matpower has been used to run the EMS applications in Matlab, including state estimation (with bad data detection) and optimal power flow algorithms. A script is implemented to exchange data with OMNeT++ scheduler over TCP/IP sockets and store measurements into a data pool. The State Estimation module uses the latest measurements from data pool to estimate the state and power flows and injections, while filtering out possible erroneous data by Bad Data Detection. For every fixed time (set to be 30 seconds), the Optimal Power Flow module uses load estimates

[2] Transmission Control Protocol (TCP) and the Internet Protocol (IP).

[3] A data table stored in a router that lists the routes of packets to particular network destinations with IP addresses.

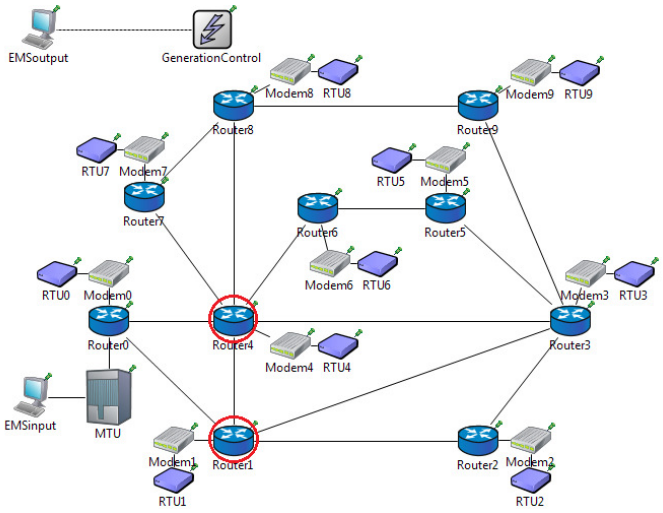


Figure 3.9: Test communication network of IEEE 14-bus system in OMNeT++.

from State Estimation to perform optimal power flow calculations and sends commands of generators' setpoints to PowerFactory through OMNeT++.

3.4.2. SIMULATORS INTEGRATION AND ATTACK MODELING

The simulation *scheduler* of OMNeT++ is customized to be the *master algorithm* for the responsibility of external integration with DIgSILENT PowerFactory and Matlab. Data is exchanged between PowerFactory, OMNeT++ and Matlab via TCP/IP sockets using the Abstract Syntax Notation One (ASN.1) protocol. On the PowerFactory side, this is implemented in the Python script that controls the simulator execution, while on the OMNeT++ side, this is implemented through the customized scheduler which adapted part of the work from [22]. This scheduler acts as the master algorithm to coordinate the simulations, handle the data exchanges with PowerFactory and Matlab, and also run OMNeT++ in a real-time mode. For the purpose of synchronization, all simulators would be started from a command after initialization and tagged with timestamps of the system clock.

An attacker can manipulate the measurements by injecting false data, making it unavailable or launching both of them. After accessing a router, the attacker can launch data integrity and availability attacks on all the data traveling through, by executing, for instance, a *man-in-the-middle attack*. By jamming, Denial of Ser-

vice (DoS) or physical attack, the attacker can block measurements in the communication links. This chapter considers the worst-case scenario that the attacker is intelligent enough with full knowledge of both the power system model and communication network model. The attacker would use the optimal combined attack policy derived from the analytic vulnerability assessment framework (3.3), i.e., try to keep stealthy from the bad data detection mechanism and manipulate the minimum number of routers and communication links. The results from the vulnerability assessment optimization program (3.3) are used to choose the routers and communication links to be attacked. These attacks are implemented in OMNeT++ by changing the behavior of the router or communication link in case it is accessed by the attacker. It should be noted that these attacks can be also modeled or simulated based on some attack modeling libraries. For instance, the NETA (NETwork Attacks) framework [108] can be used and further developed to add attack modules in the simulation model. Moreover, attacks like man-in-the-middle attack, DoS flooding attack, can be implemented by using available tools (e.g., Ettercap suite, Tribe Flood Network tool) and integrating them into the co-simulation.

3.5. NUMERICAL RESULTS

The IEEE 14-bus system in Figure 3.10 is used to perform the cyber security analysis. Mapping with Figure 3.10, let us see the difference between the abstractive communication model in Figure 3.1 for analytic vulnerability assessment and the detailed communication network in Figure 3.9 for co-simulation analysis. There are ten substations (each circle represents a substation in Figure 3.10) and the control center with SCADA master and EMS is located at the reference bus (i.e., Bus 1/node 0). There is an RTU, a modem and a router in each substation. The packets containing measurements data would be routed through multiple routers and communication links before reaching the control center. In the following the results from both analytic vulnerability assessment and the co-simulation platform are presented.

3.5.1. SECURITY INDEX UNDER THE COMMUNICATION MODEL

In order to expose vulnerability of power systems, the security index under the communication model of 14-bus system (Figure 3.1) is calculated. Note that the study of this chapter still considers the integrity and availability attacks against the SE pro-

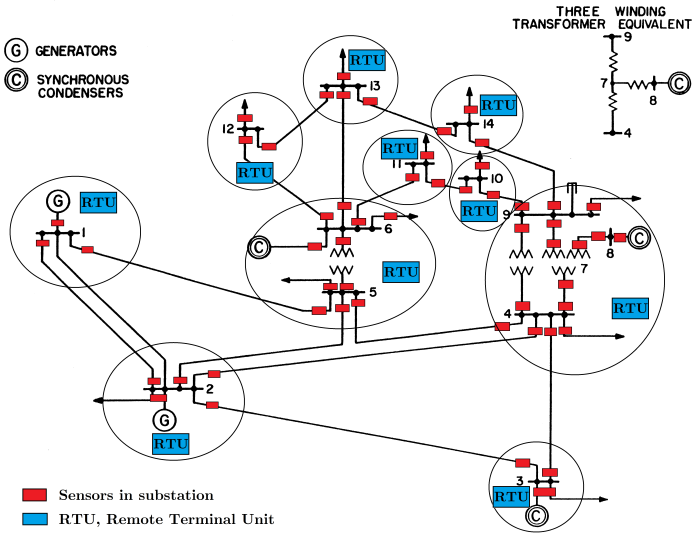


Figure 3.10: IEEE 14-bus system. There are 2 generators. Bus 1 with Generator 1 is the reference/slack bus. Generator 2 is in Bus 2. The power flow measurements are collected in each bus and both sides of the branch to provide large measurements redundancy.

cess within the EMS. The attacks here do not aim to make the system unobservable or lead to non-convergence of the SE algorithm, but instead target to introduce bias to the state estimates for further operations in the EMS. Thus only pure FDI attacks and combined (integrity and availability) attacks are considered in the security index formulation as pure availability attacks are mainly for the former purpose. For the computation, the solver CPLEX is used for the corresponding MILP optimization programs. Besides, full measurements placement is implemented in the performed experiments that power flow and injection measurements are from all the buses and transmission lines to provide large redundancy. Thus there are 54 measurements in the 14-bus system. Here the attack costs for all the measurements are calculated. For pure FDI multivariate attacks, they have to be launched on the nodes. But for combined attacks, they can take place both on nodes and communication links. It is assumed that the control center (node 0) in the 14-bus system communication model is fully protected that it can not be compromised by the attacker. To show the vulnerability of power systems to these attacks, there are no other substations or communication nodes are protected.

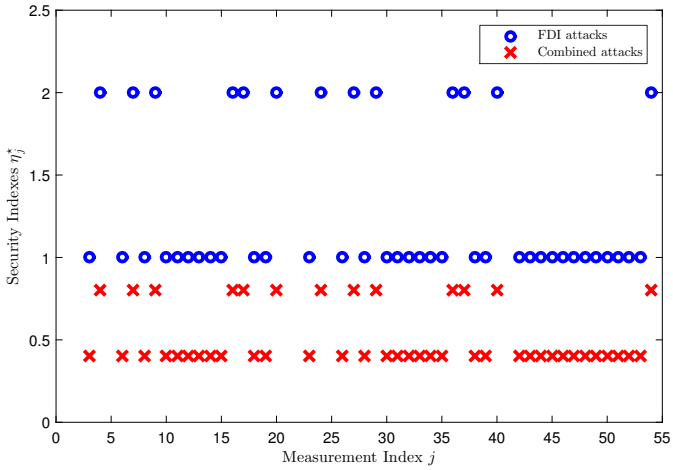


Figure 3.11: The single-path routing scheme is implemented. The security indexes $\eta_j^\star$ under combined attacks and FDI multivariate attacks are plotted versus measurement index j. Here C_V is taken as 1, and C_E is taken as 0.4.

SINGLE-PATH ROUTING

First, the single-path routing scheme is considered in the communication model, which is common in the real SCADA communication. Figure 3.11 shows the security indexes $\eta_j^\star$ of combined attacks and FDI multivariate attacks on all measurements when single-path routing scheme is implemented. As we can see, due to the protection on node 0, there are 7 measurements ($j = 1, 2, 41$ in node 0, $j = 5, 21$ in node 1 and $j = 22, 25$ in node 4) can not be attacked. Let the values become $\eta_j^\star = \infty$ for these measurements, and thus they are not shown in the Figure 3.11. Besides, the security indexes of combined attacks are smaller than the ones of pure FDI multivariate attacks. It can be inferred from Figure 3.11 that for combined attacks under single-path routing, the optimal attack strategy is to attack the integrity of the node that includes the targeted measurement and manipulate the communication links if really needed.

MULTI-PATH ROUTING

Next the multi-path routing scheme is implemented in the communication model. Here two node-disjoint routes are built for each measurement. Figure 3.12 shows results of the security indexes $\eta_j^\star$ under combined attacks and FDI attacks when it adopts multi-path routing scheme. Comparing Figure 3.11 and Figure 3.12 on the FDI attacks, it can be seen that when multi-path routing is used, it can make some

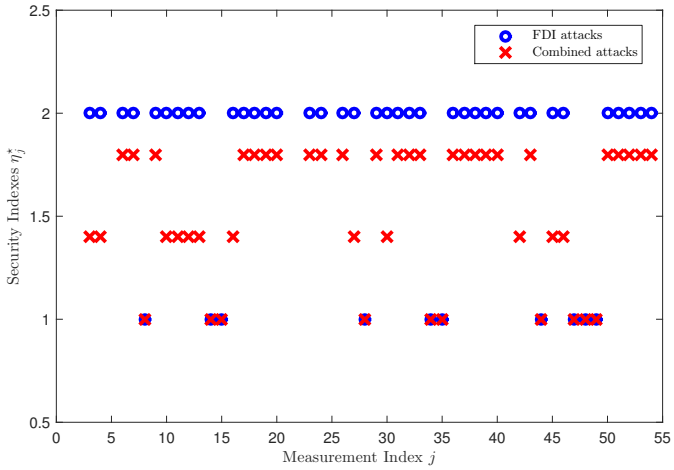


Figure 3.12: The multi-path routing scheme is implemented. The security indexes $\eta_j^{a,d}$ under combined attacks and FDI multivariate attacks are plotted versus measurement index j. C_V is taken as 1, and C_E is taken 0.4.

measurements have higher security indexes (from "1" to "2"), meaning that multi-path routing can act as a mitigation measure against pure FDI attacks. This is due to the fact that the adversary has to compromise all the routes of the measurement for FDI attacks, instead of only one route. Besides, in multi-path routing scenario, the measurements still have smaller security indexes under combined attacks. Specially from Figure 3.12, when C_V / C_E is smaller than 0.5 (0.4 is taken), all of the security indexes of combined attacks are smaller than 2. In this case, the optimal attack strategy for the adversary is to attack the integrity of the node that contains the targeted measurement and manipulate the communication links to make some other measurements unavailable.

3.5.2. CO-SIMULATION RESULTS AND DISCUSSION

In the following co-simulations are performed for cyber security analysis. The combined integrity and availability attacks have been implemented. The analytic results in Figure 3.11 and Figure 3.12 shows the minimum number of nodes and communication links to be attacked in order to corrupt specific measurements and keep stealthy. Here the single-path routing scheme is used for each measurement. According to the analytic results, node 4 and node 1 in Figure 3.1 are the most vulnerable nodes that many of the security indexes for quantifying the attacked sets

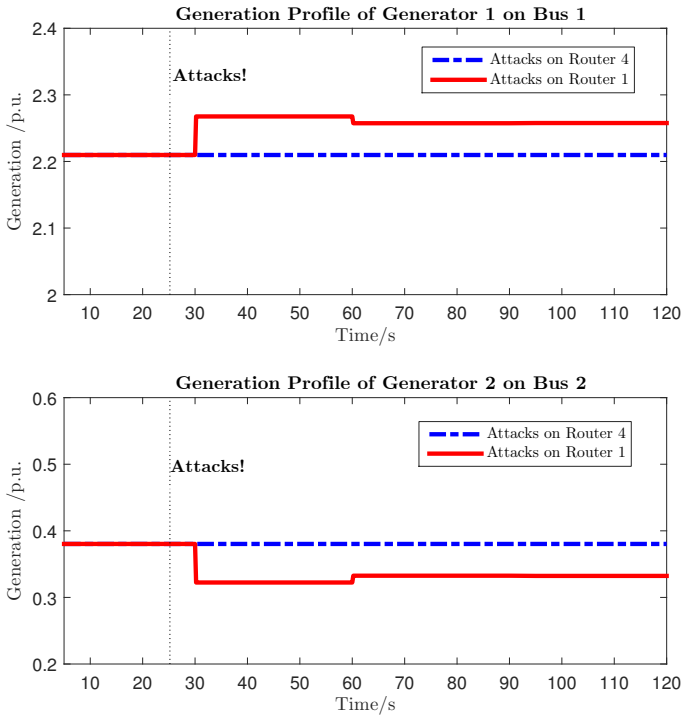


Figure 3.13: Attack impact of stealthy combined attacks on generation profile of Generator 1 and 2. The per-unit system is used and the power base is 100 MW. The true power flow measurements are generated by the simplified DC power flow calculations with Gaussian noise ($\sigma_i = 0.005$ for all the measurements). Before the occurrence of attacks, the system is operating under the optimal power flow status giving the load estimates from SE. In these two cases, the same number of measurements are corrupted.

contain these two nodes. Correspondingly, in the simulations, Router 4 (the backbone router) and Router 1 which are marked with red circles in Figure 3.9) are the most vulnerable network components. Thus in the co-simulation experiments, the "behavior" of Router 4 and Router is changed individually to simulate the attack scenarios once an attacker gains access to the internals and the packets traveling through the routers. Figure 3.13 shows the attack impact on the generation profile of generators in Bus 1 and Bus 2. Figure 3.14 shows the attack impact on the active power flows when Router 1 is attacked.

As shown in Figure 3.13, when Router 1 is attacked, the attacker misleads the system that the generation profile changes according to the setpoints. The generation power of Generator 2 has decreased and Generator 1 should compensate. The

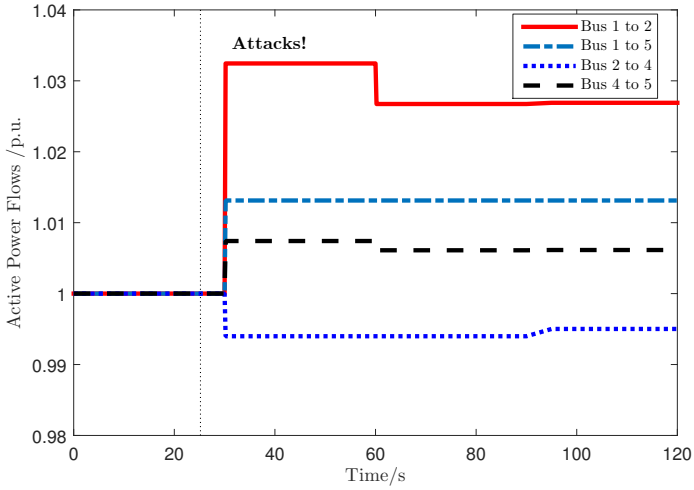


Figure 3.14: Attack impact of stealthy combined attacks on active power flows in the lines of bus 1 to 2, bus 1 to 5, bus 2 to 4, and bus 4 to 5. Router 1 is manipulated. The active power flows are normalized to the ones before the occurrence of attacks.

"latency" between the attacks' occurrence and the change of generation profile is due to the fact that the EMS sends out setpoints every 30 seconds. After the attack occurs, the generation profiles remain almost the same although the attack continues, which means the attack impact mainly depends on the initial attack magnitudes and measurements that are corrupted. When Router 4 is attacked, however, it seems that there is no attack impact on the generation profile, though Router 4 is the backbone router with the most number of packets traveling through. This is mainly because of the packets in or traveling through these two routers containing different measurements. According to the single path routing scheme, in Router 1, the attacker can gain access to the power flow measurements on bus 2, 3 and 4, which has the major impact on the generation profiles of these two generators. For the case that Router 1 is attacked, the active power flows on the transmission lines close to the generators are shown in Figure 3.14. The power flows get changed after re-dispatch according to the corrupted setpoints. Such physical impact can be utilized by the attacker to cause line overflows.

DISCUSSION ON COMBINING THEORETIC AND ICT-SPECIFIC MEASURES

The proposed analytic vulnerability assessment method can be used to explore the worst-case attack scenarios. Using the co-simulation platform, the attack impact can be evaluated by numerical simulations. Security index taking into account the physical impact of data attacks could be formulated.

As discussed in Section 3.1, co-simulation supports cyber security analysis in combining the system-theoretic and ICT-specific measures. The typical bad data detection (BDD) module acts as a theoretic measure to detect erroneous measurements. However, it fails to trigger alarms when the combined attacks on Router 1 and Router 4 are simulated since the corrupted measurements does not cause additional residual in bad data detection. To make it robust against data attacks, the communication network properties supported by co-simulation show the potentiality for developing an advanced detection scheme. For instance, when integrity attacks take place, the latency of attacked data packets changes due to the performed attack process. When availability attacks occur, the latency of attacked data packets can be treated as an extreme case. Thus a robust attack detection mechanism could be developed against combined attacks, incorporating the communication network property of the packet latency measured in the co-simulation platform. Besides, as shown in Figure 3.7, the co-simulation framework should also support the implementation of mitigation schemes such as network intrusion detection system and data authentication. For instance, the authentication scheme can be implemented by adding the configurations to the modules in OMNeT++ in the developed co-simulation platform. These ICT-specific measures together with other system-theoretic ones can be combined to propose a more advanced measure to protect intelligent power grids from cyber attacks.

4

From Static to Dynamic Detection for Power System Cyber Security

Developing advanced diagnosis tools to detect cyber attacks is crucial for power system cyber security. In the preceding chapters, it has been shown that multivariate data injection attacks can bypass bad data detection schemes typically built on static behavior of the systems, which misleads operators to disruptive decisions. This chapter would depart from the existing static viewpoint to develop a diagnosis filter that captures the dynamics signatures of such a multivariate intrusion. To this end, this chapter introduces a dynamic residual generator approach formulated as robust optimization programs in order to detect a class of disruptive multivariate attacks that potentially remain stealthy in view of a static bad data detector. Two possible desired features are investigated: (i) a non-zero transient and (ii) a non-zero steady-state behavior of the residual generator in the presence of an attack. In case (i), the

This chapter is based on the following published work:
[109] K. Pan, P. Palensky, & P. Mohajerin Esfahani (2019). From Static to Dynamic Anomaly Detection with Application to Power System Cyber Security. *IEEE Transactions on Power Systems*, pp, 1–1. DOI: 10.1109/TPWRS.2019.2943304.

problem is reformulated as a finite, but possibly non-convex, optimization program. A linear programming relaxation is further developed, which improves the scalability, and as such practicality, of the diagnosis filter design. In case (ii), it turns out that the resulting robust program admits an exact convex reformulation, yielding a Nash equilibrium between the attacker and the residual generator. This assertion has an interesting implication: the proposed approach is not conservative in the sense that the additional knowledge of the worst-case attack does not improve the diagnosis performance. To illustrate the theoretical results, the proposed diagnosis filter is implemented to detect multivariate attacks on the system measurements deployed to generate the so-called Automatic Generation Control signals in a three-area IEEE 39-bus system.

4.1. INTRODUCTION

4.1.1. BACKGROUND AND RELATED WORK

As discussed in Section 1.1.2, National Institute of Standards and Technology (NIST) [26] defines five functions for protecting Information and Communication Technology (ICT): (i) Identify, (ii) Protect, (iii) Detect, (iv) Respond, (v) Recover. It would be naive to think an ICT system can be sufficiently protected in order to address the issues raised by (iii)-(v). From this chapter the thesis starts to focus on (iii) Detection of false data injection (FDI) attacks for supervisory control and data acquisition (SCADA) systems, which are in charge of transmitting measurement and control signals between power system substations and control centers [110]. Such SCADA systems are notorious for being based on legacy ICT, and are a popular target for adversaries [1, 111] nowadays. The consequences of a successful attack on SCADA systems can be catastrophic to an economy and society in general [11, 112]. In this light, it is of utmost importance to detect attacks and respond accordingly.

Literature on anomaly detection Traditionally, SCADA systems deploy bad data detection (BDD) to filter out possible erroneous measurements due to sensor failures or anomalies [17]. The BDD process captures only a snapshot of the steady states of system trajectories, and thus only exploits possible *static* impact of intrusions. Although this method can perform successfully in detecting basic attacks, it may fail in the presence of the so-called *stealthy multivariate attacks* that carefully launch synthesized false data injections given full knowledge of the system model [14]; recall the FDI attacks introduced in Chapter 2.

The work in [13] first shows that such an attack can corrupt the state estimation function without being detected by BDD. Since then vulnerability and impact analysis of stealthy attacks on power systems have been a prominent subject in the literature. As thoroughly analyzed in Chapter 2, a notion to quantify the vulnerability to stealthy attacks, so-called security index, is directly concerned with the level of efforts required to alter specific measurements [35, 113]. Without advanced diagnosis tools, tampering measurements remains undetected, causing state deviations, equipment damages or even cascading failures [45]. Techniques proposed to deal with stealthy attacks include statistical methods such as sequential detection using Cumulative Sum (CUSUM)-type algorithms [30], and measurements consistency

assessment under certain observability assumptions [31]. A detection method that leverages online information is described in [32], which is applicable by ensuring the availability and accuracy of load forecasts and generation schedules. In [114], a mechanism is introduced to formulate the detection scheme as a matrix separation problem, but it only recovers intrusions among corrupted measurements over a particular period of time. These techniques are essentially static detection methods that may be confined by certain prior assumptions on the distribution of measurement errors. Despite an extensive and ongoing literature focusing on the static part of BDD mechanism, the following question remains largely unexplored:

Would it be possible to detect stealthy multivariate attacks in a real-time operation by exploiting the attack impact on the dynamics of system trajectories during the transient behavior?

The importance of an appropriate answer to this question has been reinforced thanks to recent advances in sensing technology in the modern power systems. The main objective of this chapter is to address this question.

Related work Detection approaches concerning system dynamics have primarily emerged under the topic of *fault detection and isolation filters*. A subclass of these schemes is the observer-based approach applied initially to linear models [115]; see also [116] for a comprehensive summary of the large body of literature. The authors in [117] further extend the modeling framework to general linear differential-algebraic equations (DAEs), enhancing the applicability of such methods particularly for power system applications due to the common governing physical laws in this setting. Recently, a variant of observer-based methods is also investigated in [118] so as to deal with unknown natural exogenous inputs.

An inherent shortcoming of many observer-based approaches is that the degree of the resulting diagnosis filter is effectively the same as the system dynamics, which may yield an unnecessarily complex filter in large-scale power systems. To the best knowledge of this thesis, there are relatively much fewer studies in the literature on the design of the reduced-order observers where the conditions for a minimum order existence need to be satisfied [116, 119]. The closest approach in the literature is [120] where a scalable optimization-based filter design is developed for high-dimensional nonlinear control systems. However, the proposed method opts

for mainly dealing with a single fault scenario, and may not be as effective in case of smart multivariate adversarial inputs.

An effective approach toward security and modeling the interaction between attackers and detectors builds on the rich framework of game theory. Recently, the authors in [121] propose a two player mixed strategy game to address a dynamic resource-planning problem between an attacker targeting the communication equipment and a defender protecting the control network. Similar frameworks have also been deployed to model the dynamics of information flow between an advanced persistent threat and a detector [122, 123].

4.1.2. CONTRIBUTIONS AND OUTLINE

The main objective of this chapter is to develop a *diagnosis filter* to detect *FDI multivariate attacks* in a real-time operation. For this purpose, considering a class of disruptive multivariate attack scenarios (Definition 4.2.5), the effects of attacks on power system dynamics can be characterized through a set of differential equations. Having transferred the dynamics into the discrete-time domain, the diagnosis filter is further restricted to a family of dynamic residual generators that entirely decouples the contributions of the attacks from the system states and natural disturbances. In order to identify an admissible multivariate attack scenario, an optimization-based framework is proposed to robustify the diagnosis filter with respect to such attacks, i.e., aiming to design a filter whose residual (output) is sensitive to any plausible disruptive multivariate attacks. The main contributions are:

(i) Unlike some existing work, this chapter goes beyond a static viewpoint of attack detection to capture the attack impact on the dynamics of system trajectories. The diagnosis filter design approach is characterized as a robust optimization program. It is guaranteed that while the filter residual is decoupled from system states and disturbances, it still remains sensitive to admissible disruptive multivariate attacks even if the attacker has full knowledge about the diagnosis filter architecture (Definition 4.4.1 and the program (4.18)).

(ii) To detect attacks during the transient behavior, the resulting robust program is reformulated as a finite, possibly non-convex, optimization program (Theorem 4.4.3). To improve the scalability of the proposed solution, a linear pro-

gramming relaxation which is highly tractable for large scale systems (Corollary 4.4.4) is proposed. It is guaranteed that if the optimal value of the relaxed program is positive, the resulting diagnosis filter is able to detect any admissible disruptive attack scenarios, which may remain stealthy through the lens of a static detector.

(iii) This chapter work further explores the steady-state behavior of the diagnosis filter (Lemma 4.4.6). In this case, an exact convex reformulation of the resulting robust program is developed. As a byproduct, it is shown that the proposed solution is indeed a Nash equilibrium between the attacker and the residual generator (Theorem 4.4.7). An interesting implication of such a Nash equilibrium is that the information of the attack signal may not necessarily improve the performance of the diagnosis filter. In other words, if the proposed convex optimization fails to have a desirable feasible solution, then there may exist a disruptive stealthy attack in the long-term horizon where the exact knowledge of the attack signal does not help in designing a successful residual generator.

In addition to the above theoretical results, the effectiveness of the proposed diagnosis filter is validated on a multi-area IEEE 39-bus system. Numerical results illustrate that the diagnosis filter successfully generates a residual "alert" in the presence of multivariate attacks that are stealthy in a static viewpoint, even in a noisy environment with imprecise measurements.

Section 4.2 introduces the problem of power system cyber security, and the challenges posed by stealthy multivariate attacks are highlighted. Section 4.3 discusses a model instance of power system dynamics under attacks on measurements. The diagnosis filter design is proposed in Section 4.4 where an optimization framework is introduced, and numerical simulations are reported in Section 4.5.

4.2. PROBLEM STATEMENT: EFFECTS OF ATTACKS ON SYSTEM DYNAMICS

4.2.1. STATIC DETECTION AND SYSTEM MODELING

From the chapter this thesis work moves from the description of system in steady-state towards system dynamics. As discussed in the preceding, for a power grid,

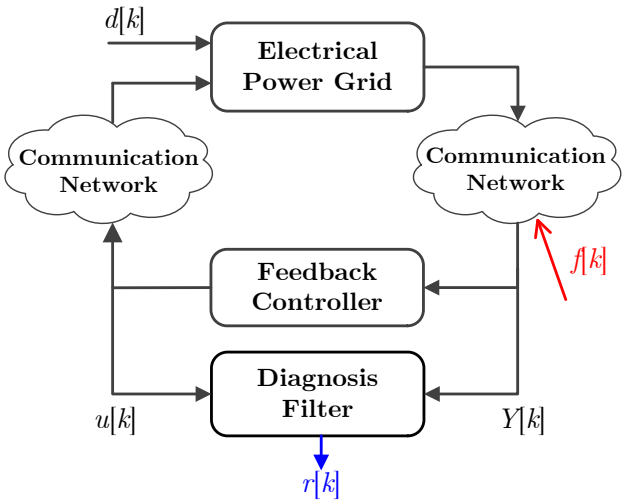


Figure 4.1: Schematic block diagram of the general system model.

measurements are collected by remote sensors in the substations and transmitted through a SCADA communication network. The typical BDD is conducted to detect the erroneous measurements at each time instance. This can be treated as a static process: it only concerns the system states $X[k] \in \mathbb{R}^{n_X}$ and measurements $Y[k] \in \mathbb{R}^{n_Y}$ at time step $k \in \mathbb{N}$, which can be described by

$$Y[k] = CX[k] + D_f f[k], \tag{4.1}$$

where $C \in \mathbb{R}^{n_Y \times n_X}$ is the measurement matrix, and $f[\cdot] \in \mathbb{R}^{n_f}$ represents the false data injection attacks on measurements. Different from the expression in (2.1), here we add a timestamp for these variable to highlight their trajectories in a time horizon. The matrix D_f is to characterize which measurement is vulnerable to attacks. It is customary to define a *residual signal* for a static detector, $r_S[k] := Y[k] - \hat{Y}[k]$, where $\hat{Y}[\cdot]$ denotes the estimated measurements. As we know, in the traditional weighted least squares estimation, the estimate of state is $(C^\top C)^{-1} C^\top Y[k]$, assuming that C has full column rank with high measurement redundancy. Then the measurements estimate is $C(C^\top C)^{-1} C^\top Y[k]$, and the residual signal can be further expressed as

$$r_S[k] = \left(I_{n^Y} - C(C^\top C)^{-1} C^\top\right) Y[k]. \tag{4.2}$$

Such an anomaly detector has shown a good effectiveness in detecting erroneous data and basic attacks [124]. However, in the face of coordinated attacks on

multiple measurements, this static detector can fail. Motivated by this shortcoming, this chapter work takes a dynamic design perspective where it shifts the emphasis on an attack as a static process to the attack's effects on power system dynamics. In particular, it opts for differentiating the attack impact on the systems trajectories from natural unknown disturbances such as load deviations.

To model its impact on the system dynamics, let us consider a more general modeling framework in Figure 4.1. The electrical grid is operated by a digital controller that receives measurements as inputs and sends control signals to the actuators through SCADA communication networks. These transmitted data are applied in discrete-time samples. On the power grid side, the input $d[k] \in \mathbb{R}^{n_d}$ represents natural disturbances. On the controller side, a control signal $u[k] \in \mathbb{R}^{n_u}$ is computed given the measurements $Y[k]$. Note that with the closed-loop control, the corruptions $f[k]$ on the measurements would affect the system dynamics. The dynamics of the closed-loop system is

$$\begin{cases} X[k+1] = A_x X[k] + B_d d[k] + B_u u[k], \\ Y[k] = CX[k] + D_f f[k], \end{cases} \tag{4.3}$$

where A_x, B_d and B_u are constant matrices. Let us highlight the difference between the dynamical system (4.3) and the respective static counterpart (4.1). In fact, the time independence of the first equation in (4.3) describes the dynamics of the system, while the algebraic equation (4.1) represents the relation on each time instance and describes a static relation between the states and outputs. The aim of this study is to exploit such dynamics information in (4.3) in order to design a diagnosis filter to detect stealthy multivariate attacks. To illustrate the attack impact on the system dynamics, one can simply consider the feedback controller as a linear operator such that $u[k] = GY[k]$ where $G \in \mathbb{R}^{n_u \times n_Y}$ is a matrix gain. By defining the closed-loop system matrices $A_{cl} := A_x + B_u GC$ and $B_f := B_u G D_f$, one can reformulate (4.3) into

$$\begin{cases} X[k+1] = A_{cl} X[k] + B_d d[k] + B_f f[k], \\ Y[k] = CX[k] + D_f f[k]. \end{cases} \tag{4.4}$$

Remark 4.2.1 (Dynamic feedback controller)**.** *The restriction to only a static feedback controller $u[k] = GY[k]$ to transfer from (4.3) to (4.4) is without loss of generality. Namely, the proposed framework is rich enough to subsume a dynamic controller architecture as well. Indeed, when the controller has certain dynamics, it suffices to*

augment the system dynamics (4.3) with the controller states and outputs. It is referred to Section 5.2.1 of Chapter 5 for such a detailed analysis.

Remark 4.2.2 (Attacks impact on the dynamics of system trajectories)**.** *In light of (4.4), matrices B_f, D_f capture the attack impact on the power system dynamics, mapping attacks $f[\cdot]$ to the system states and measurements respectively.*

The following description shows that the state-space description (4.4) is a particular case of DAE model. By introducing a time-shift operator $q: qX[k] \rightarrow X[k+1]$, one can fit (4.4) into

$$H(q)x[k] + L(q)y[k] + F(q)f[k] = 0, \tag{4.5}$$

where $x := [X^\top \ d^\top]^\top$ represents the unknown signals of system states and disturbances, $y := Y$ contains all the available data for the operator. Let n_x and n_y be the dimensions of $x[\cdot]$, $y[\cdot]$, and let n_r be the number of rows in (4.5). Then H, L, F are polynomial matrices in terms of the time-shift operator q with n_r rows and n_x, n_y, n_f columns separately, by defining,

$$H(q) := \begin{bmatrix} -qI + A_{cl} & B_d \\ C & 0 \end{bmatrix}, \quad L(q) := \begin{bmatrix} 0 \\ -I_{n^Y} \end{bmatrix}, \quad F(q) := \begin{bmatrix} B_f \\ D_f \end{bmatrix}.$$

4.2.2. CHALLENGE: STEALTHY MULTIVARIATE ATTACKS

This subsection starts with an existing result characterizing the set of stealthy multivariate attacks that can bypass the static detector.

Lemma 4.2.3 (Stealthy attack values [13, Theorem 1])**.** *Consider the measurement equation (4.1) and the static detector with the respective residual function (4.2). Then, an attack $f[\cdot]$ remains stealthy, i.e., it does not cause any additional residue to (4.2), if it takes values from the set*

$$\mathcal{F} := \Big\{ f[k] \in \mathbb{R}^{n_f} : \ D_f f[k] \in Im(C), \quad k \in \mathbb{N} \Big\}. \tag{4.6}$$

One can observe that a stealthy attack $D_f f[\cdot]$ described in (4.6) has the knowledge of the system model (4.1) through the range space of C. That is, it represents a tampered value $D_f f[k] = C\Delta X$ where $\Delta X \in \mathbb{R}^{n_X}$ can be any injected bias influencing certain sensor measurements. Such multivariate attacks would also challenge the detector design as they may neutralize the outputs of the diagnosis filter.

Assumption 4.2.4 (Stationary attacks)**.** *Throughout this chapter, the study mainly considers attacks $f[\cdot]$ that are time-invariant, i.e., $f[k] = 0$ for all $k \leq k_{\min}$; $f[k] = f \in \mathcal{F}$ for all $k > k_{\min}$. Namely, the attack occurs as a constant bias injection f on measurements during the system operations at a specific unknown time instance $k_{\min}$, and it remains unchanged since then.*

Advanced attacks also pursue a maximized impact on the system dynamics. Thus, an adversary would try to inject "smart" false data, possibly with large magnitudes, in such a way that it causes the maximum damage. The next definition opts to formalize this class of attacks.

Definition 4.2.5 (Disruptive stealthy attack)**.** *Consider a matrix $F_{\mathrm{b}} := [f_1, f_2, \ldots, f_d]$ representing a finite basis for the set of stealthy attacks (4.6), i.e., the set $\mathcal{F}$ defined in (4.6) can equivalently be represented by*

$$\mathcal{F} = \left\{ F_{\mathrm{b}}^{\top} \alpha = \sum_{i=1}^{d} \alpha_i f_i \;\middle|\; \alpha = [\alpha_1, \alpha_2, \cdots, \alpha_d]^{\top} \in \mathbb{R}^d \right\}.$$

A signal $f \in \mathcal{F}$ is called disruptive stealthy attack *if its corresponding coefficients α is a polytopic set, i.e., it belongs to*

$$\mathcal{A} := \left\{ \alpha \in \mathbb{R}^d \mid A\alpha \geq b \right\}, \tag{4.7}$$

where $A \in \mathbb{R}^{n_b \times d}$ and $b \in \mathbb{R}^{n_b}$ are given matrices. It is emphasized that the subsequent analysis and the proposed diagnosis filter design only rely on the convexity of the set $\mathcal{A}$. Namely, the choice (4.7) may be adjusted according to the application at hand, as long as the convexity of the set is respected.

4.3. MODELING INSTANCE OF POWER SYSTEM DYNAMICS

4.3.1. STATE-SPACE MODEL OF ONE-AREA AGC SYSTEM

This section would go through a modeling instance of power system dynamics in the form of (4.4): Automatic Generation Control (AGC) closed-loop system under attacks. This model will be used to validate the designed diagnosis filter. Figure 4.2 depicts the diagram of a three-area IEEE 39-bus system. AGC is a feedback controller that tunes the setpoints of participated generators (e.g., G11 of Area 1) to maintain the frequency as its nominal value and the tie-line (e.g., L1-2 between Area 1 and 2) power as the scheduled one.

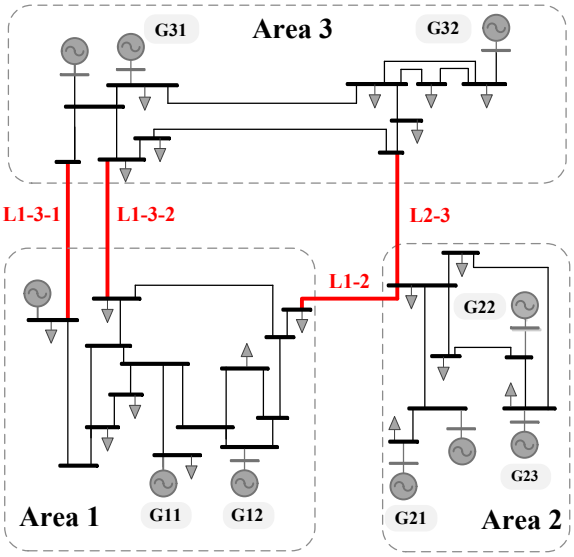


Figure 4.2: Three-area 39-bus system: the measurements of the tie-lines (in red) L1-3, L1-2, L2-3 are attacked.

In the work of AGC, a linearized model is commonly used for the load frequency dynamics [125]. For a three-area system in Figure 4.2, the frequency dynamics in Area i can be written as

$$\Delta\dot{\omega}_i = \frac{1}{2H_i}\left(\Delta P_{m_i} - \Delta P_{tie_i} - \Delta P_{l_i} - D_i\Delta\omega_i\right), \tag{4.8a}$$

where H_i is the equivalent inertia constant, D_i is the damping coefficient and ΔP_{l_i} denotes load deviations. Here ΔP_{tie_i}, ΔP_{m_i} represent the total tie-line power exchanges from Area i and the total generated power in Area i, i.e., $\Delta P_{tie_i} = \sum_{j\in\mathcal{E}_i}\Delta P_{tie_{i,j}}$ where $\mathcal{E}_i$ denotes the set of areas that connect to Area i, and $\Delta P_{m_i} = \sum_{g=1}^{G_i}\Delta P_{m_{i,g}}$ where G_i denotes the number of participated generators in Area i. One can have

$$\Delta\dot{P}_{m_{i,g}} = -\frac{1}{T_{ch_{i,g}}}\left(\Delta P_{m_{i,g}} + \frac{1}{S_{i,g}}\Delta\omega_i - \phi_{i,g}\Delta P_{agc_i}\right), \tag{4.8b}$$

$$\Delta\dot{P}_{tie_{i,j}} = T_{ij}\left(\Delta\omega_i - \Delta\omega_j\right), \tag{4.8c}$$

where $T_{ch_{i,g}}$ is the governor-turbine's time constant, $S_{i,g}$ denote the droop coefficient, T_{ij} is the synchronizing parameter between Area i and j. Note that ΔP_{agc_i} is the signal from AGC for the participated generators to track the load changes, and $\phi_{i,g}$ is the participating factor, i.e., $\sum_{g=1}^{G_i}\phi_{i,g} = 1$. After receiving the frequency and

tie-line power measurements, the *area control error* (ACE) is computed for an integral action in the AGC system,

$$ACE_i = B_i \Delta\omega_i + \sum_{j\in\mathcal{E}_i} \Delta P_{tie_{i,j}}, \tag{4.8d}$$

$$\Delta\dot{P}_{agc_i} = -K_{I_i} ACE_i, \tag{4.8e}$$

where B_i is the frequency bias and K_{I_i} represents the integral gain. Based on the equations (4.8), the linearized model of Area i can be presented as the state equation

$$\dot{X}_i(t) = A_{ii}X(t) + B_{i,d}d_i(t) + \sum_{j\in\mathcal{E}_i} A_{ij}X_j(t), \tag{4.9}$$

4

where X_i is the state vector, $d_i := \Delta P_{l_i}$ denotes load deviations. Recall Remark 4.2.1 that (4.9) is an augmented model for the closed-loop AGC system that X_i consists of not only the electrical grid states (e.g., frequency, generator output and tie-line power) but also the controller state ΔP_{agc_i}, i.e.,

$$X_i := \left[\{\Delta P_{tie_{i,j}}\}_{j\in\mathcal{E}_i} \quad \Delta\omega_i \quad \{\Delta P_{m_{i,g}}\}_{1:G_i} \quad \Delta P_{agc_i}\right]^\top.$$

Besides in (4.9), A_{ii} is the system matrix of Area i, A_{ij} is a matrix whose only non-zero element is $-T_{ij}$ in row 1 or 2 and column 3, and $B_{i,d}$ is the matrix for load deviations.

In addition to (4.9), a measurement model with high redundancy is assumed that the measurements of each tie-line power ($\Delta P_{tie_{i,j}}$) and the total tie-lines' power (ΔP_{tie_i}), the frequency ($\Delta\omega_i$), each generator output ($\Delta P_{m_{i,g}}$) and the total generated power (ΔP_{m_i}), and the AGC controller output (ΔP_{agc_i}) are all available. Besides, vulnerabilities within SCADA networks may allow cyber intrusions. Thus the system output equation is

$$Y_i(t) = C_i X(t) + D_{i,f} f_i(t), \tag{4.10}$$

where Y_i is the system output and C_i is the output tall-matrix with full column rank. Here f_i denotes multivariate attacks and the matrix $D_{i,f}$ quantifies which output is attacked. In the aforementioned section, due to the feedback loop, attacks on the measurements would also affect the frequency dynamics. Hence the state equation (4.9) during attacks becomes

$$\dot{X}_i(t) = A_{ii}X(t) + B_{i,d}d_i(t) + B_{i,f}f_i(t) + \sum_{j\in\mathcal{E}_i} A_{ij}X_j(t),$$

where $B_{i,f}$ is the matrix that relates attacks to system states.

4.3.2. STATE-SPACE MODEL OF MULTI-AREA AGC SYSTEM

Using the state equations of each area, the continuous-time model of the three-area system can be obtained,

$$\dot{X}(t) = \tilde{A}_{cl} X(t) + \tilde{B}_d d(t) + \tilde{B}_f f(t), \tag{4.11}$$

where X is the vector consisting of groups of system dynamic states in each area, d is the vector for all areas' load deviations, and f denotes all the attack signals in the three-area, namely,

$$X = \begin{bmatrix} X_1^\top & X_2^\top & X_3^\top \end{bmatrix}^\top, \; d = \begin{bmatrix} \Delta P_{l_1} & \Delta P_{l_2} & \Delta P_{l_3} \end{bmatrix}^\top, \; f = \begin{bmatrix} f_l^\top & f_2^\top & f_3^\top \end{bmatrix}^\top.$$

In (4.11), $\tilde{A}_{cl}$ is the closed-loop system matrix, $\tilde{B}_d$, $\tilde{B}_f$ are constant matrices that relate load deviations and attacks to system states. For the three-area system, these matrices are

$$\tilde{A}_{cl} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}, \; \tilde{B}_d = \text{diag}\begin{bmatrix} B_{1,d}, & B_{2,d}, & B_{3,d} \end{bmatrix}, \; \tilde{B}_f = \text{diag}\begin{bmatrix} B_{1,f}, & B_{2,f}, & B_{3,f} \end{bmatrix}.$$

One can also obtain the output equation of the system,

$$Y(t) = CX(t) + D_f f(t), \tag{4.12}$$

where Y is the system output vector containing all the three areas' outputs, C is the output matrix, and D_f quantifies all the vulnerable signals. These matrices are

$$Y = \begin{bmatrix} Y_1^\top & Y_2^\top & Y_3^\top \end{bmatrix}^\top, \; C = \text{diag}\begin{bmatrix} C_1, & C_2, & C_3 \end{bmatrix}, \; D_f = \text{diag}\begin{bmatrix} D_{1,f}, & D_{2,f}, & D_{3,f} \end{bmatrix}.$$

To obtain the sampled discrete-time model as (4.4), (4.11) and (4.12) must be discretized. A zero-order hold (ZOH)[1] discretization is deployed [126],

$$A_{cl} = e^{\tilde{A}_{cl} T_s}, \quad B_d = \int_0^{T_s} e^{\tilde{A}_{cl}(T_s - t)} \tilde{B}_d \, \mathrm{d}t. \tag{4.13}$$

where T_s is the sampling time. Note that the attack matrix $\tilde{B}_f$ has the same matrix transformation as $\tilde{B}_d$, resulting B_f. The above approximation is exact for a ZOH and (4.13) corresponds to the analytical solution of the discretization. Therefore, the above model can be described in the form of (4.4) which again can be fitted into the DAE (4.5). In Appendix 4.7, a detailed description of the involved parameters of the three-area system as well as the attack scenarios is provided.

[1] The inputs signals $d(\cdot)$ and $f(\cdot)$ in (4.11) are assumed to be piecewise constant within the sampling periods.

4.4. ROBUST DYNAMIC DETECTION

4.4.1. PRELIMINARIES FOR DIAGNOSIS FILTER CONSTRUCTION

An ideal detection aims to implement a non-zero mapping from the attack to the diagnostic signal while decoupled from system states and disturbances, given the available data $y[\cdot]$ in the control center. In the power system dynamics described via a set of DAE, the diagnosis filter is restricted to a type of dynamic residual generator in the form of linear transfer functions, i.e., $r_D[k] := R(q)y[k]$ where r_D is the residual signal of the diagnosis filter and $R(q)$ is a transfer operator. Note that $y[\cdot]$ is associated with the polynomial matrix $L(q)$ in (4.5). A formulation of the transform operator $R(q)$ can be

$$R(q) := a(q)^{-1}N(q)L(q),$$

where $N(q)$ is a polynomial vector with the dimension of n_r and a predefined order d_N. To make $R(q)$ physically realizable, stable dynamics $a(q)$ with sufficient order need to be added as the denominator where all the roots are strictly contained in the unit circle. Note that, unlike the observer-based methods, d_N can be much less than the dimension of system dynamics. Then $N(q)$ and $a(q)$ are the two variables for the filter design. By multiplying $a(q)^{-1}N(q)$ in the left of (4.5), one can have

$$\begin{aligned} r_D[k] &= a(q)^{-1}N(q)L(q)y[k] \\ &= -\underbrace{a(q)^{-1}N(q)H(q)x[k]}_{\text{(I)}} - \underbrace{a(q)^{-1}N(q)F(q)f[k]}_{\text{(II)}}, \end{aligned} \tag{4.14}$$

where term (I) in (4.14) is due to $x[\cdot]$ of system states and natural disturbances. Term (II) is the desired contribution from the attacks $f[\cdot]$. In view of this diagnosis filter description, a class of residual generator which is sensitive to disruptive stealthy attacks as defined in Definition 4.2.5 can be introduced.

Definition 4.4.1 (Robust residual generator)**.** *Consider a linear residual generator represented via a polynomial vector $N(q)$. This residual generator is robust with respect to disruptive stealthy attacks introduced in Definition 4.2.5 if*

$$\begin{cases} (I) & N(q)H(q) = 0, \\ (II) & N(q)F(q)F_{\mathrm{b}}\alpha \neq 0, \quad \forall \alpha \in \mathcal{A}, \end{cases} \tag{4.15}$$

where the basis matrix F_{b} and the set $\mathcal{A}$ are the same as the ones in Definition 4.2.5.

In the next step, it would be shown that the polynomial equations (4.15) in Definition 4.4.1 can be characterized as a feasibility problem of a finite robust program.

Lemma 4.4.2 (Linear program characterization)**.** *Consider the polynomial matrices* $H(q)=\sum_{i=0}^{1}H_i q^i$, $N(q):=\sum_{i=0}^{d_N}N_i q^i$ *and* $F(q)=F$, *where* $H_i\in\mathbb{R}^{n_r\times n_x}$, $N_i\in\mathbb{R}^{n_r}$, *and* $F\in\mathbb{R}^{n_r\times n_f}$ *are constant matrices. Then, the family of robust residual generators in* (4.15) *is characterized by*

$$\begin{cases} (I) & \bar{N}\bar{H}=0, \\ (II) & \left\|\bar{N}V(\alpha)\right\|_\infty>0, \quad \forall\alpha\in\mathcal{A}, \end{cases} \tag{4.16}$$

where $\|\cdot\|_\infty$ *is the infinity vector norm, and* $\bar{N}:=\begin{bmatrix}N_0 & N_1 & \cdots & N_{d_N}\end{bmatrix}$,

$$\bar{H}:=\begin{bmatrix}H_0 & H_1 & 0 & \cdots & 0\\ 0 & H_0 & H_1 & 0 & \vdots\\ \vdots & 0 & \ddots & \ddots & 0\\ 0 & \cdots & 0 & H_0 & H_1\end{bmatrix},\quad V(\alpha):=\begin{bmatrix}FF_{\mathrm{b}}\alpha & 0 & \cdots & 0\\ 0 & FF_{\mathrm{b}}\alpha & 0 & \vdots\\ \vdots & 0 & \ddots & 0\\ 0 & \cdots & 0 & FF_{\mathrm{b}}\alpha\end{bmatrix}.$$

Proof. The proof follows a similar line of arguments as [120, Lemma 4.2]. The key step is to observe that $N(q)H(q)=\bar{N}\bar{H}[I_{n^x},\ qI_{n^x},\ \cdots,\ q^{d_N+1}I_{n^x}]^\top$, and also it satisfies $N(q)FF_{\mathrm{b}}\alpha=\bar{N}V(\alpha)[I_d,\ qI_d,\ \cdots,\ q^{d_N}I_d]^\top$. The rest of the proof follows rather straightforwardly, and the details are omitted for the sake of brevity. ■

4.4.2. ROBUST DIAGNOSIS FILTER: TRANSIENT BEHAVIOR

In light of (4.16), one can define a symmetric set for the design variable $\bar{N}$ of the dynamic residual generator,

$$\mathcal{N}:=\left\{\bar{N}\in\mathbb{R}^{(d_N+1)n_r}\mid \bar{N}\bar{H}=0,\ \|\bar{N}\|_\infty\le\eta\right\}. \tag{4.17}$$

The second constraint in the set is added to avoid possible unbounded solutions. To design a robust residual generator, the goal is to find an $\bar{N}\in\mathcal{N}$ that for all $\alpha\in\mathcal{A}$, (4.16) can be satisfied. A natural reformulation of the residual synthesis is to consider an objective function as the second quantity in (4.16) influenced by the parameters $\mathcal{N}$ and the attacker action α, i.e., $\mathcal{J}(\bar{N},\alpha):=\|\bar{N}V(\alpha)\|_\infty$. A successful scenario from an attacker viewpoint is to minimize this objective function given a residual generator. Therefore, this study takes a rather conservative viewpoint where the attacker may have complete knowledge of the system model and even the residual

generator parameters, and exploits it so as to synthesize a stealthy attack. Such a conservative view even takes into account the worst-case scenario where a powerful attacker (maybe an internal attacker or a sponsored attacker) can access all the corresponding modules even the repressive measures in the system operations. Then the diagnosis filter design can be reformulated as the robust program,

$$\gamma^\star := \max_{\bar{N}\in\mathcal{N}} \min_{\alpha\in\mathcal{A}} \left\{ \mathcal{J}(\bar{N},\alpha) := \|\bar{N}V(\alpha)\|_\infty \right\}. \tag{4.18}$$

The optimal value $\gamma^\star$ of the robust reformulation (4.18) is indeed an indication whether the attack still remains stealthy in the dynamic setting, i.e., if $\gamma^\star > 0$ then the optimal solution $\bar{N}^\star$ yields a diagnosis filter in the form of (4.14) which detects all the admissible attacks introduced in Definition 4.2.5. However, if $\gamma^\star = 0$, then it implies that for any possible detectors (static or dynamic) there exists a stationary disruptive attack that remains stealthy. In the next step, it is shown that the robust program (4.18) can be equivalently reformulated as a finite (non-convex) optimization problem.

Theorem 4.4.3 (Finite reformulation of (4.18))**.** *The robust optimization* (4.18) *can be equivalently described via the finite optimization program*

$$\begin{aligned} \gamma^\star = \max_{\bar{N},\,\beta,\,\lambda}\ & b^\top\lambda \\ s.t.\ & \sum_{i=0}^{d_N} \left(\beta_{2i}-\beta_{2i+1}\right) N_i F F_{\mathrm{b}} = \lambda^\top A, \\ & \mathbf{1}^\top\beta = 1,\ \beta \geq 0, \\ & \bar{N}\in\mathcal{N},\ \lambda \geq 0, \end{aligned} \tag{4.19}$$

where $\beta = \left[\beta_0,\ \beta_1,\ \cdots,\ \beta_{2d_N+1}\right]^\top$ *is an* $\mathbb{R}^{2d_N+2}$*-valued auxiliary variable.*

Proof. See Appendix 4.6.1. ■

The exact reformulation program (4.19) for (4.18) is unfortunately non-convex due to the bilinearity between the variables β and N_i in the first constraint. In the following corollary, a convex relaxation of the program is proposed by restricting the feasible set of the variable β to a $2d_N+2$ finite possibilities where $\beta = \left[0,\cdots,1,\cdots,0\right]^\top$ in which the only non-zero element of the vector is the i-th element.

Corollary 4.4.4 (Linear program relaxation)**.** *Given $i \in \{1, \ldots, 2d_N+2\}$, consider the linear program*

$$\begin{aligned} \gamma_i^\star := \quad & \max_{\bar{N},\lambda} \quad b^\top \lambda \\ & s.t. \quad (-1)^i N_{\lfloor i/2 \rfloor} F F_{\mathrm{b}} = \lambda^\top A, \\ & \quad\;\; \bar{N} \in \mathcal{N},\ \lambda \geq 0, \end{aligned} \tag{LP$_i$}$$

where $\lfloor \cdot \rfloor$ is the ceiling function that maps the argument to the least integer. Then, the solution to the program (LP$_i$) is a feasible solution to the exact robust design reformulation (4.19), and $\max_{\{i \leq 2d_N+2\}} \gamma_i^\star \leq \gamma^\star$. In particular, if for any $i \in \{1, \ldots, 2d_N+2\}$ one can have $\gamma_i^\star > 0$, then the solution to program LP$_i$ offers a robust residual generator detecting all admissible disruptive attacks introduced by Definition 4.2.5.

Corollary 4.4.4 suggests that the maximum optimal value of $\{\gamma_0^\star, \gamma_1^\star, \cdots, \gamma_{2d_N+2}^\star\}$ and its corresponding $\bar{N}^\star$ provide a suboptimal solution to the original robust design (4.18).

It should be noted that the focus of this chapter is on stationary (time-invariant) attacks. It is also important to highlight that the robust design perspective (4.18) allows the attacker to know the system model and filter parameters. In such a setting, the detection procedure could be much more difficult if the attacker would be able to dynamically adapt the attack values over the time, i.e., the attack signal is time-varying. In fact, in a multivariate attack scenario, one can construct a disruptive time-varying attack bypassing any linear residual generators. The next remark alludes more to this situation.

Remark 4.4.5 (Time-varying stealthy attacks)**.** *Consider a multivariate attack $f = \begin{bmatrix} f_1 & f_2 & \cdots & f_{n_f} \end{bmatrix}^\top$ where each element is a time-varying signal $f_i = f_i[k]$. Then, the residual (4.14) can be rewritten as*

$$a(q) r_D[k] = -\sum_{i=1}^{n_f} \Big(N(q) F_i f_i[\cdot] \Big)[k], \tag{4.20}$$

where $F = \begin{bmatrix} F_1 & F_2 & \cdots & F_{n_f} \end{bmatrix}$ represents the attack dynamics matrix. One can inspect that when the time-varying relation $\sum_{i=1}^{n_f} \big(N(q) F_i f_i[\cdot] \big)[k] = 0$ holds for every k, for instance when

$$f_{n_f}[k] = -\big(N(q) F_{n_f} \big)^{-1} \sum_{i=1}^{n_f-1} \Big(N(q) F_i f_i[\cdot] \Big)[k],$$

then the residual outcome (4.20) *stays zero for all k, and as such, the attack remains undetected for any linear residual generator.*

The proposed robust design in (4.18) does not necessarily enforce a non-zero steady-state residual of the diagnosis filter under multivariate attacks. Namely, the design perspective of (4.18) focuses on detection of attacks during the transient behavior without any requirements on long-term behavior of the residual. Indeed, the residual signal r_D may return to zero value after a successful reaction to the attack's occurrence. A more stringent perspective is to require a non-zero steady-state behavior of the diagnosis filter under any admissible attack scenario in $\alpha \in \mathcal{A}$. This extension is addressed in the next subsection.

4.4.3. ROBUST DIAGNOSIS FILTER: STEADY-STATE BEHAVIOR

In order to design a diagnosis filter with non-zero steady-state residual "alert" when a multivariate attack occurs, the robust optimization (4.18) can be modified by a more conservative (smaller) objective function $\mathcal{J}(\bar{N}, \alpha) := |\bar{N}\bar{F}\alpha|$ where

$$\bar{F} := \begin{bmatrix} FF_{\mathrm{b}} & FF_{\mathrm{b}} & \cdots & FF_{\mathrm{b}} \end{bmatrix}^{\top}. \tag{4.21}$$

A similar treatment as the preceding subsection can establish a framework for computational purposes. The next lemma follows similar objective as in Lemma 4.4.2 with a more demanding requirement of the non-zero long-term residual behavior.

Lemma 4.4.6 (Non-zero steady-state residual characterization)**.** *For the polynomial matrices $H(q)$, $N(q)$ and $F(q)$ as defined in Lemma 4.4.2, the family of dynamic residual generators with non-zero steady-state residual under multivariate attacks can be characterized by the algebraic relations*

$$\begin{cases} (I) & \bar{N}\bar{H} = 0, \\ (II) & |\bar{N}\bar{F}\alpha| > 0, \quad \forall \alpha \in \mathcal{A}, \end{cases} \tag{4.22}$$

where $\bar{F}$ is defined in (4.21)*, and the matrices $\bar{N}, \bar{H}$ are as defined in Lemma 4.4.2.*

Proof. Recall that $N(q)H(q) = \bar{N}\bar{H}\begin{bmatrix} I_{n^x}, & qI_{n^x}, & \cdots, & q^{d_N+1}I_{n^x} \end{bmatrix}^{\top}$. Thus if $\bar{N}\bar{H} = 0$, the diagnosis filter becomes $r_D[k] = -a(q)^{-1}N(q)f[k]$. Note the steady-state value of the filter residual under attacks would be $-a(q)^{-1}N(q)F(q)f|_{q=1}$. Thus for the multivariate attack with the coefficient α, the steady-state value of the filter residual is $-a(1)^{-1}N(1)F(1)F_{\mathrm{b}}\alpha$. The proof concludes by noting that $N(1)F(1)F_{\mathrm{b}}\alpha = \bar{N}\bar{F}\alpha$. ■

In a similar fashion, the robust design perspective in (4.18) can be modified accordingly as

$$\mu^\star := \max_{\bar{N}\in\mathcal{N}} \min_{\alpha\in\mathcal{A}} \Big\{\mathcal{J}(\bar{N},\alpha) := |\bar{N}\bar{F}\alpha|\Big\}. \tag{4.23}$$

Notice the relation between the new objective function with the absolute value and the one in (4.18) with the infinity-norm. As it appears in the next result, the new setting is in fact a restricted case of the finite reformulation in Theorem 4.4.3.

Theorem 4.4.7 (Residual long-term behavior: exact convex reformulation and Nash equilibrium)**.** *Consider the minimax counterpart of the program* (4.18) *as defined*

$$\varphi^\star := \min_{\alpha\in\mathcal{A}} \max_{\bar{N}\in\mathcal{N}} \Big\{\mathcal{J}(\bar{N},\alpha) := |\bar{N}\bar{F}\alpha|\Big\}. \tag{4.24}$$

Each of the program (4.23) *and* (4.24) *can be equivalently reformulated through the linear programs*

$$\begin{aligned} \mu^\star = \quad & \max_{\bar{N},\,\lambda} && b^\top\lambda \\ & s.t. && \bar{N}\bar{F} = \lambda^\top A \\ & && \bar{N}\in\mathcal{N},\ \lambda\geq 0, \end{aligned} \tag{4.25a}$$

$$\begin{aligned} \varphi^\star = \quad & \min_{v_1,v_2,w,\alpha} && \mathbf{1}^\top v_1 + \mathbf{1}^\top v_2 \\ & s.t. && \bar{H}w + v_1 - v_2 = \bar{F}\alpha \\ & && v_1\geq 0,\ v_2\geq 0, \\ & && A\alpha\geq b. \end{aligned} \tag{4.25b}$$

Moreover, the value of each of these two programs coincide, i.e., $\mu^\star = \varphi^\star$.

Proof. See Appendix 4.6.2. ■

It is worth noting the difference between the robust perspective of (4.23) versus the minimax program (4.24). While in the design perspective of (4.23) the filter is oblivious to the possible attack scenarios, in the perspective of (4.24) the filter is aware of the attack signal and opts to detect that particular signal in the presence of natural disturbances. Obviously, the former setting is the one closer to the reality and, in general, the knowledge of the attack signal should help the detection significantly. This observation can indeed be translated through the usual weak inequality of $\mu^\star \leq \varphi^\star$. However, Theorem 4.4.7 indicates that the filter performance,

in view of the long-term behavior of the worst-case attack scenario, indeed does not depend on the exact knowledge of the attacker signal and the inequality holds as the equality. This discussion is summarized in the following remark.

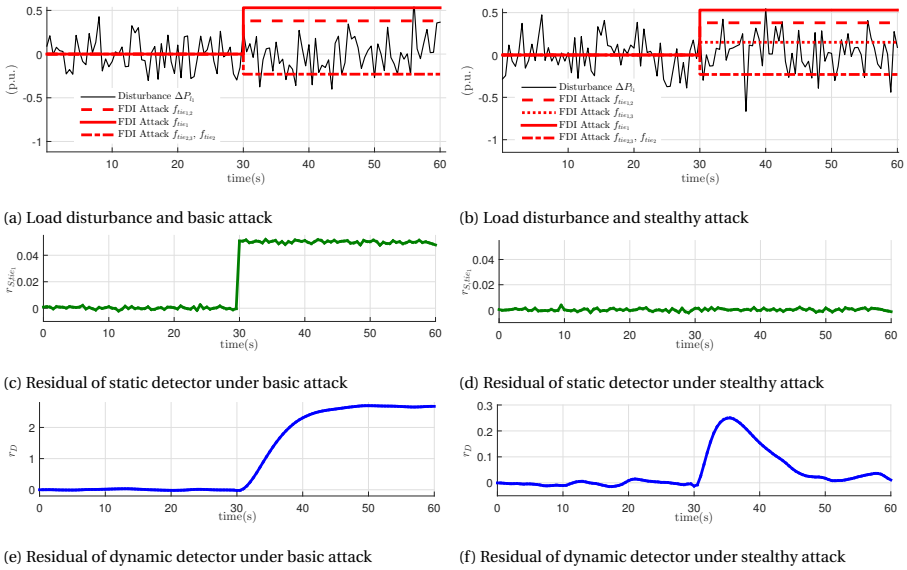
Remark 4.4.8 (Nash equilibrium interpretation)**.** *If the linear programs* (4.25a) (4.25b) *admit a positive optimal value* $\varphi^\star = \mu^\star > 0$*, then the resulting filter can detect all the admissible multivariate attacks described by Definition* 4.2.5 *along with a non-zero steady-state residual level. On the other hand, if the optimal values coincide with* $\varphi^\star = \mu^\star = 0$*, it then implies that there is no linear filter being able to decouple the admissible attack with* $\alpha^\star$*, the solution to* (4.25b)*, from the natural disturbances in a long-term horizon.*

4.5. NUMERICAL RESULTS

4.5.1. TEST SYSTEM AND DIAGNOSIS FILTER DESCRIPTION

In order to validate the effectiveness of the diagnosis filter with application to power system cyber security, the IEEE 39-bus system which is well-known as a standard system for testing of new power system analysis is used. As shown in Figure 4.2, this system consists of 3 areas and 10 generators where 7 of them are equipped with AGC for frequency control. All the participating generators in each area are with equal participation factors. The total load of the three-area system is 5.483 GW for the base of 100 MVA and 60 Hz. The generator specifications and AGC parameters of each area are referred to the work [127], and the linear frequency dynamics model has been developed in the preceding Section 4.3. Thus finally it results in a 19-order model in the form of (4.4).

The diagnosis filter proposed in Section 4.4 is applied to detect the multivariate disruptive stealthy attacks on the measurements of the AGC system. In the following simulations, the degree of the dynamic residual generator is set to $d_N = 3$ which is much less than the order of the dynamics model. A sampling time $T_s = 0.5\,\text{s}$ and a finite time horizon 60 s are chosen for all simulations. To design the filter, the denominator is set to have the form $a(q) = (q-p)^{d_N}/(1-p)^{d_N}$ where p is a user-defined variable acting as the *pole* of the transfer operator $R(q)$, and it is normalized in steady-state value for all feasible poles. The pole is set to $p = 0.8$ for a stable dynamic behavior at the beginning the simulations, and the solver CPLEX has been



(a) Load disturbance and basic attack

(b) Load disturbance and stealthy attack

(c) Residual of static detector under basic attack

(d) Residual of static detector under stealthy attack

(e) Residual of dynamic detector under basic attack

(f) Residual of dynamic detector under stealthy attack

Figure 4.3: Static detector in (4.2) versus dynamic detector (diagnosis filter) from Corollary 4.4.4 under basic and stealthy attacks.

used to solve the corresponding optimization problems.

4.5.2. SIMULATION RESULTS

To evaluate the performance of the diagnosis filter, the disturbances $d_i = \Delta P_{l_i}$ are modeled as stochastic load patterns [2] to capture its uncertainty. As shown in Figure 4.3a and Figure 4.3b, the load deviation ΔP_{l_1} in Area 1 is modeled as random zero-mean Gaussian signals. It should be noted that tie-line power flow measurements are much more vulnerable to cyber attacks, comparing with frequency measurements (e.g., the anomalies in frequency can be easily detected by comparing the corrupted reading with the normal one.) [129]. Therefore as indicated in Figure 4.2 this study mainly focuses on the scenario that there are 5 vulnerable tie-line power measurements, i.e., $\Delta P_{tie_{1,2}}$, $\Delta P_{tie_{1,3}}$, ΔP_{tie_1}, $\Delta P_{tie_{2,3}}$ and ΔP_{tie_2}. Recalling Definition 4.2.5 for stealthy attack basis, thus there exist 3 basis vectors in the spanning set and they can be modeled as follows: $f_1 = \begin{bmatrix}0.1 & 0 & 0.1 & 0 & 0\end{bmatrix}^T$, $f_2 = \begin{bmatrix}0.1 & 0.15 & 0.25 & 0 & 0\end{bmatrix}^T$, $f_3 = \begin{bmatrix}0 & 0 & 0 & 0.1 & 0.1\end{bmatrix}^T$ (all in p.u.). Here each basis vector lies in the range space of the output matrix that the corrupted measurements still align with an actual physical

[2] High frequency load fluctuation is typically time uncorrected stochastic noise on second time scale [128].

state, bypassing the static detector. Furthermore, without loss of generality, the parameters are set to $A = \mathbf{1}^\top$ and $b = 1.5$ in the set $\mathcal{A}$ and $\eta = 10$ in the set $\mathcal{N}$. The design variable $\bar{N}$ of the robust residual generator is first derived by solving (4.18) through (LP_i). The optimal value achieves maximum for $i = 2$ that $\gamma_2^\star = 300$, which implies a robust detection during the transient behavior as Corollary 4.4.4. For the given $\bar{N}$, the multivariate attack coordinates $\alpha = \begin{bmatrix}2.8 & 1 & -2.3\end{bmatrix}^\top$ are obtained by solving the inner minimization of (4.18). Next, the steady-state behavior of the filter with the above sets $\mathcal{N}$ and $\mathcal{A}$ is studied. To do that, following Theorem 4.4.7 the programs (4.23) and (4.24) are solved through the resulted programs (4.25a) and (4.25b). It turns out that the derived optimal values satisfy the equality $\varphi^\star = \mu^\star = 0$, indicating that the optimal multivariate attack with $\alpha^\star$, the optimizer of the program (4.25b) and an optimal solution to (4.24), is a stealthy attack in the long-term horizon. To highlight, thanks to the fact that the optimal values of the programs (4.25a) (4.25b) form a Nash equilibrium, even with the exact information of the stealthy attack coefficients $\alpha^\star$, the diagnosis filter still cannot decouple the long-term behavior of the residual from the natural disturbances; see Remark 4.4.8.

The first simulation begins with a general scenario where the multivariate attack is not carefully coordinated, i.e., basic attack. Thus as shown in Figure 4.3a, only 4 of 5 vulnerable measurements are compromised that $f_{tie_{1,2}} = 0.38\,\mathrm{p.u.}$, $f_{tie_1} = 0.53\,\mathrm{p.u.}$, $f_{tie_{2,3}} = -0.23\,\mathrm{p.u.}$ and $f_{tie_2} = -0.23\,\mathrm{p.u.}$. Note that since the injected data on $\Delta P_{tie_{1,2}}$ and ΔP_{tie_1} are inconsistent, the static detector is also expected to be triggered. To test the detectors in a more realistic setup, the presence of process and measurements noises is also considered. The process noise term added to the state equation of Area 1 is zero-mean Gaussian noises with the covariance matrix $R_{X_1} = 0.03 \times \mathrm{diag}\big([1\ \ 1\ \ 0.03\ \ 1\ \ 1\ \ 1\ \ 1]^\top\big)$, i.e., the covariance of the noise to the frequency is 0.009 and the covariance of other states' noise is 0.03 [118]. Similarly, the measurement noise term added to the measurements of Area 1 is with the covariance matrix $R_{Y_1} = 0.03 \times \mathrm{diag}\big([1\ \ 1\ \ 1\ \ 0.03\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1]^\top\big)$, i.e., the covariance of the frequency measurement is 0.009 and the covariance of other measurements' noise is 0.03 [118]. Note the residue r_S of BDD in (4.2) becomes $r_S[k] = (I - C(C^\top R_Y^{-1} C)^{-1} C^\top R_Y^{-1}) Y[k]$ under the noisy system. The attacks are launched at $k_{\min} = 30\,\mathrm{s}$. In Figure 4.3c and Figure 4.3e, results of the static detector in (4.2) and the proposed dynamic detector (diagnosis filter) are presented. Both detectors have succeeded to generate a

diagnostic signal when attacks occurred, and the diagnosis filter residual r_D is significantly decoupled from stochastic load disturbances, and keeps sensitive to the multivariate attacks for a successful detection under noisy system settings.

In the second simulation, to challenge the detectors, now the multivariate attacks have been launched on all the 5 vulnerable measurements and the derived attack coefficient α from the optimization results has been used for a more intelligent adversary. Thus in Figure 4.3b, the corruptions become $f_{tie_{1,2}} = 0.38\,\text{p.u.}$, $f_{tie_{1,3}} = 0.15\,\text{p.u.}$, $f_{tie_1} = 0.53\,\text{p.u.}$, $f_{tie_{2,3}} = -0.23\,\text{p.u.}$ and $f_{tie_2} = -0.23\,\text{p.u.}$. This corresponds to the worst case for the diagnosis filter that the adversary is given the knowledge of the residual generator's parameter $\bar{N}$ that it tries to minimize the payoff function over $\mathcal{A}$. Besides, the noisy system settings have been considered. Figure 4.3d and Figure 4.3f demonstrate all the simulation results. In Figure 4.3d, the static detector becomes totally blind to the occurrence of such an intelligent attack. However, as we can see in Figure 4.3f, even in the worst case, the diagnosis filter works perfectly well under the noisy system, generating a residual "alert" for the presence of multivariate attacks. We can also see that the residual output becomes close to zero again after a successful detection during the transient behavior in Figure 4.3f, which is consistent to the aforementioned result $\varphi^\star = \mu^\star = 0$ and Remark 4.4.8. These simulations prove the effectiveness and robustness of the proposed diagnosis filter design.

4.5.3. FURTHER DISCUSSIONS

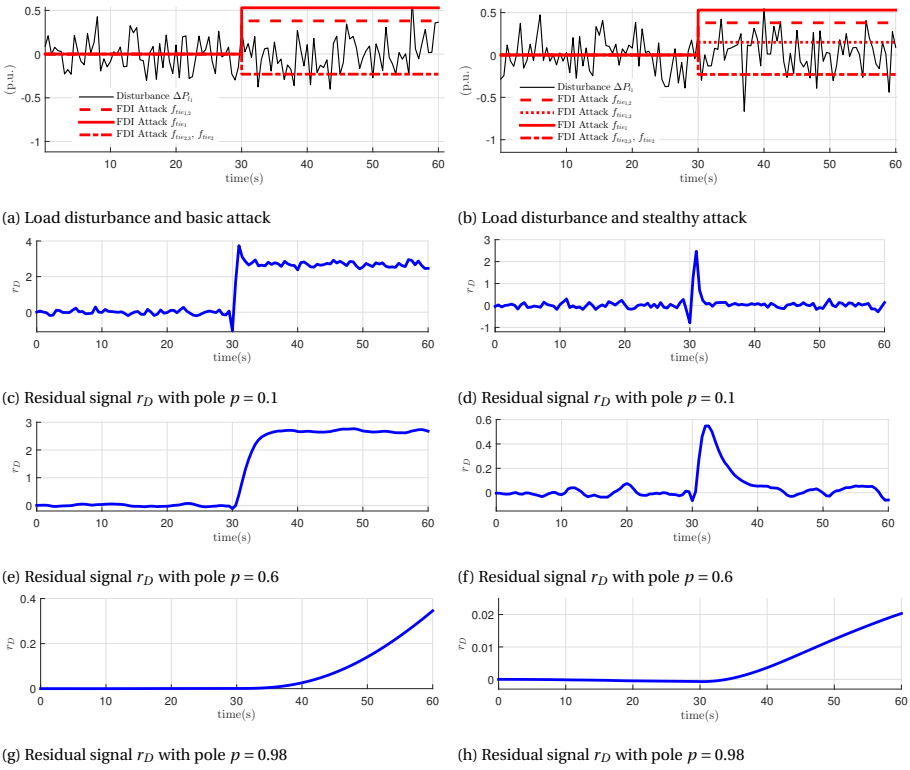
This section elaborates several practical aspects of the proposed filter in the preceding sections.

DIAGNOSIS SENSITIVITY TO FILTER POLES

While the denominator of the filter $a(q)$ in (4.14) is chosen rather arbitrarily, up to a stability condition, the poles however has a significant impact on the residual sensitivity. As a general rule, the smaller the poles, the faster the residual responds, and the more sensitive the residual responds to model imprecision and noises. Simulation results in Figure 4.4 numerically illustrate this relation when the poles vary.

OTHER TYPES OF ATTACKS

In addition to the multivariate measurement attack, the main focus of this study, there are several other types of attacks that are briefly discussed in the following:



(a) Load disturbance and basic attack

(b) Load disturbance and stealthy attack

(c) Residual signal r_D with pole $p = 0.1$

(d) Residual signal r_D with pole $p = 0.1$

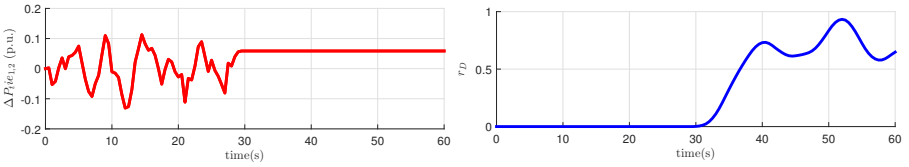
(e) Residual signal r_D with pole $p = 0.6$

(f) Residual signal r_D with pole $p = 0.6$

(g) Residual signal r_D with pole $p = 0.98$

(h) Residual signal r_D with pole $p = 0.98$

Figure 4.4: Results of dynamic detector (diagnosis filter) with different *poles* ($p = 0.1,\ 0.6,\ 0.98$) under basic and stealthy attacks.



(a) $\Delta P_{tie_{1,2}}$ under DoS attacks from $k_{dos} = 30$

(b) Residual signal r_D under DoS attacks

Figure 4.5: Results of dynamic detector (diagnosis filter) under DoS attacks on $\Delta P_{tie_{1,2}}$ ($p = 0.8$).

- *Denial-of-service (DoS) attack*: A type of availability attack where the attacker aims to prevent some specific data from being delivered to the respective destinations.

- *Replay attack*: A two-stage attack where the adversary gathers a sequence of data packets at stage 1, and replays the recorded data afterwards at stage 2.

To recall, the DoS attack scenario in the context of combined attacks is introduced and discussed in Chapter 2. Though no additional alerts would the triggered in the current BDD for DoS attacks, from a detection point of view, they are still trivially detectable without any sophisticated mechanisms as the absence of data is not stealthy. Thus a specific missing data detection scheme has been proposed for such a validation in Section 2.6.4 of Chapter 2. Besides, in the typical DoS attack modeling, the missing data is replaced with the last received ones [130]. In such a mechanism, the DoS can be treated as an "injection" attack. The performance of the developed filter in the presence of this class of attacks is investigated in Figure 4.5. Numerical results confirm that the proposed filter can successfully detect the DoS attacks. In regard with the replay attack, the articles [131, 132] offer sufficient conditions under which plausible attacks may remain stealthy irrespective of the detection mechanism if the attacker has accessed all the necessary data and excite the attack of stage 2 at a suitable time.

OBSERVER-BASED DIAGNOSIS FILTERS

Another major technique for attack detection builds on observer-based techniques. In this view, the estimate of the system states, or in more general setting *output observer*, is a reference to alert the abnormality [133]. This section is closed by a brief summary of the differences between these approaches and the one of this study.

- The observer-based approaches typically yield diagnosis filters with higher dynamical system degrees than the approach proposed in this chapter. A low-order diagnosis filter is often more desired due to practical aspects of online implementation particularly for large-scale power systems.
- Observer-based diagnosis filters usually rely on a precondition of system observability. An extended version of such observer-based filters relaxes this condition to the so-called Luenberger-type conditions [134]. The proposed diagnosis filter, however, requires a weaker condition reflected through the feasibility condition of the resulting optimization programs, e.g., when the program (4.16) in Lemma 4.4.2 is feasible.
- Thanks to the optimization-based framework, unlike the observer-based approaches, the study of this chapter has a systematic approach to incorporate a multivariate attack scenario into the framework.

4.6. APPENDIX I: TECHNICAL PROOFS

4.6.1. PROOF OF THEOREM 4.4.3

Let us recall that $\bar{N}V(\alpha) = \begin{bmatrix} N_0 F F_{\mathrm{b}} \alpha & N_1 F F_{\mathrm{b}} \alpha & \cdots & N_{d_N} F F_{\mathrm{b}} \alpha \end{bmatrix}$, and as such, the payoff function of the robust reformulation (4.18) is $\mathcal{J}(\bar{N}, \alpha) = \max_i |N_i F F_{\mathrm{b}} \alpha|$ where $i \in \{0, \cdots, d_N\}$. By introducing an auxiliary variable β in the simplex set $\mathcal{B} := \{\beta \in \mathbb{R}^{2d_N+2} \mid \beta \geq 0,\ \mathbf{1}^\top \beta = 1\}$, one can rewrite $\mathcal{J}$ as

$$\mathcal{J}(\bar{N}, \alpha) = \max_{\beta \in \mathcal{B}} \sum_{i=0}^{d_N} (\beta_{2i} - \beta_{2i+1}) N_i F F_{\mathrm{b}}.$$

In this light, the original robust strategy (4.18) can be equivalently described via

$$\max_{\bar{N} \in \mathcal{N}} \min_{\alpha \in \mathcal{A}} \max_{\beta \in \mathcal{B}} \left\{ \sum_{i=0}^{d_N} (\beta_{2i} - \beta_{2i+1}) N_i F F_{\mathrm{b}} \alpha \right\}.$$

Note that given a fixed $\bar{N}$ the inner minimax optimization is indeed a bilinear objective in the decision variables and the respective feasible sets $\mathcal{A}$ and $\mathcal{B}$ are convex. Since one of the sets, $\mathcal{B}$, is also compact, then the zero-duality gap holds. Therefore, interchanging the optimization over $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$ yields

$$\gamma^\star = \max_{\bar{N} \in \mathcal{N},\ \beta \in \mathcal{B}} \left\{ \min_{\alpha \in \mathcal{A}} \sum_{i=0}^{d_N} (\beta_{2i} - \beta_{2i+1}) N_i F F_{\mathrm{b}} \alpha \right\}. \tag{4.26}$$

The inner minimization of (4.26) is a (feasible) linear program. The duality can be used again. To this end, let us assume that the decision variables $\bar{N}$ and β are fixed and consider the Lagrangian function

$$\mathcal{L}(\alpha; \lambda) = b^\top \lambda + \Big(\sum_{i=0}^{d_N} (\beta_{2i} - \beta_{2i+1}) N_i F F_{\mathrm{b}} - \lambda^\top A \Big) \alpha,$$

where optimizing over an unconstrained variable α becomes

$$\min_{\alpha} \mathcal{L}(\alpha; \lambda) = \begin{cases} b^\top \lambda & \text{if } \begin{cases} \sum_{i=0}^{d_N} (\beta_{2i} - \beta_{2i+1}) N_i F F_{\mathrm{b}} = \lambda^\top A \\ \lambda \geq 0 \end{cases} \\ -\infty & \text{otherwise,} \end{cases}$$

Using the above characterization as the most inner optimization program in (4.26)

would lead to the following program,

$$\begin{aligned} \max_{\lambda} \quad & b^\top \lambda \\ \text{s.t.} \quad & \sum_{i=0}^{d_N} \big(\beta_{2i} - \beta_{2i+1}\big) N_i F F_{\mathrm{b}} = \lambda^\top A, \\ & \lambda \geq 0. \end{aligned} \tag{4.27}$$

It then suffices to combine maximizing over the auxiliary variable λ together with the variables $\bar{N}$ and β to arrive at the main result in (4.19).

4.6.2. PROOF OF THEOREM 4.4.7

Let us first prove the convex reformulation. For a given $\bar{N} \in \mathcal{N}$, the inner minimization of (4.23) can be translated as

$$\begin{aligned} \min_{\alpha \in \mathcal{A},\, r} \quad & r \\ \text{s.t.} \quad & \bar{N}\bar{F}\alpha - r \leq 0, \\ & -\bar{N}\bar{F}\alpha - r \leq 0. \end{aligned}$$

The Lagrangian of the inner minimization reads as

$$\mathcal{L}(\alpha, r; \beta, \lambda) = b^\top \lambda + \Big(\big(\beta_0 - \beta_1\big)\bar{N}\bar{F} - \lambda^\top A\Big)\alpha + \big(1 - \beta_0 - \beta_1\big) r.$$

Optimizing over the variables $\alpha,\ r$ yields

$$\min_{\alpha,\, r} \mathcal{L}(\alpha, r; \beta, \lambda) = \begin{cases} b^\top \lambda & \text{if} \begin{cases} \big(\beta_0 - \beta_1\big)\bar{N}\bar{F} = \lambda^\top A \\ \beta_0 + \beta_1 \leq 1 \\ \beta_0 \geq 0,\ \beta_1 \geq 0,\ \lambda \geq 0 \end{cases} \\ -\infty & \text{otherwise.} \end{cases}$$

Then, combining maximization over the auxiliary variables λ, β_0, β_1 together with the variable $\bar{N}$ arrives at the optimization program,

$$\begin{aligned} \mu^\star = \quad \max_{\bar{N}, \beta_0, \beta_1, \lambda} \quad & b^\top \lambda \\ \text{s.t.} \quad & \big(\beta_0 - \beta_1\big)\bar{N}\bar{F} = \lambda^\top A, \\ & \beta_0 + \beta_1 \leq 1,\ \beta_0 \geq 0,\ \beta_1 \geq 0, \\ & \bar{N} \in \mathcal{N},\ \lambda \in \mathbb{R}^{n_b},\ \lambda \geq 0. \end{aligned} \tag{4.28}$$

Note that the actual program (4.25a) is a restriction of (4.28) where the variables β_0 and β_1 are restricted to $\beta_0 = 1$ and $\beta_1 = 0$. Next, it is shown that this restriction is

indeed without loss of generality. To this end, suppose the tuple $(\beta_0^\star, \beta_1^\star, \bar{N}^\star, \lambda^\star)$ is an optimal solution to the program (4.28). Note that the optimal variables $\beta_0^\star$ and $\beta_1^\star$ may satisfy one of the following three properties:

(i) $\beta_0^\star = \beta_1^\star$: In this case, $\lambda^\star = 0$, and therefore the optimal value $\mu^\star = 0$. This optimal solution can be trivially achieved in the program (4.25a) by setting $\bar{N} = 0$.

(ii) $\beta_0^\star > \beta_1^\star$: Observe that the tuple $\big(\beta_0^{'} = 1,\ \beta_1^{'} = 0,\ \bar{N}^{'} = \bar{N}^\star,\ \lambda^{'} = \lambda^\star/(\beta_0^\star - \beta_1^\star)\big)$ is a feasible solution with the objective value $b^\top\lambda^\star/(\beta_0^\star - \beta_1^\star)$. Since $b^\top\lambda^\star \geq 0$ by optimality assumption and $\beta_0^\star - \beta_1^\star \in (0, 1]$, then this feasible solution has a possibly higher optimal value, and therefore $\beta_0^\star - \beta_1^\star = 1$. That is, $\beta_0^\star = 1$ and $\beta_1^\star = 0$.

(iii) $\beta_0^\star < \beta_1^\star$: Following similar steps as the previous case together with the symmetric property of the feasible set $\mathcal{N}$, one can show that the optimal value of the program (4.28) also coincides with the restricted version in (4.25a).

This concludes the proof of the convex reformulation from (4.23) to (4.25a). In regard with the minimax problem (4.24), let us recall the symmetric property of the feasible set $\mathcal{N}$ in the variable $\bar{N}$. With a fixed α, the inner maximization can be directly formed as $\max_{\bar{N}\in\mathcal{N}} \bar{N}\bar{F}\alpha$ whose Lagrangian becomes

$$\mathcal{L}(\bar{N}; v, w) = -\big(\mathbf{1}^\top v_1 + \mathbf{1}^\top v_2\big) + \big(w^\top \bar{H}^\top + v_1^\top - v_2^\top - (\bar{F}\alpha)^\top\big)\bar{N}^\top,$$

Optimizing over the variable $\bar{N}$ leads to

$$\min_{\bar{N}} \mathcal{L}(\bar{N}; v, w) = \begin{cases} -\mathbf{1}^\top v_1 - \mathbf{1}^\top v_2 & \text{if } \begin{cases} \bar{H}w + v_1 - v_2 = \bar{F}\alpha \\ v_1 \geq 0,\ v_2 \geq 0 \end{cases} \\ -\infty & \text{otherwise.} \end{cases}$$

Thus, combining minimization over the auxiliary variables $v_1,\ v_2,\ w$ together with the variable α, the minimax optimization (4.24) can be reformulated as the linear program (4.25b).

Finally, It would show that the solution to programs (4.25) indeed forms a Nash equilibrium between the programs (4.23) and (4.24). Till now, the maximin and minimax problems are reformulated as linear programs (4.25). The idea is to show that

these programs have the same optimal values. In fact, it can be shown that the programs are dual of each other, and that the strong duality holds when both programs are feasible. To this end, one can resort to the duality of (4.25a) with the Lagrangian

$$\mathcal{L}(\bar{N},\lambda;\alpha,v,w)=\left(w^\top\bar{H}^\top+v_1^\top-v_2^\top-(\bar{F}\alpha)^\top\right)\bar{N}^\top+\left(\alpha^\top A^\top-b^\top\right)\lambda-\left(\mathbf{1}^\top v_1+\mathbf{1}^\top v_2\right).$$

Optimizing over the variables $\bar{N}$, λ yields

$$\min_{\bar{N},\lambda}\mathcal{L}(\bar{N},\lambda;\alpha,v,w)=\begin{cases}-\mathbf{1}^\top v_1-\mathbf{1}^\top v_2 & \text{if}\begin{cases}\bar{H}w+v_1-v_2=\bar{F}\alpha\\ A\alpha\geq b\\ v_1\geq 0,\ v_2\geq 0\end{cases}\\ -\infty & \text{otherwise.}\end{cases}$$

It is not difficult to see that the above program coincides with the program (4.25b); this concludes the proof.

4.7. APPENDIX II: SYSTEM PARAMETERS

In this section, the involved matrices and parameters of the three-area 39 system are provided. Let us take the model description of Area 1 in the three-area system in Figure 4.2 of Section 4.3 as an instance,

$$B_{1,d}=\begin{bmatrix}0 & 0 & -\frac{1}{2H_1} & 0 & 0 & 0\end{bmatrix}^\top,$$

$$A_{11}=\begin{bmatrix}0 & 0 & T_{12} & 0 & 0 & 0\\ 0 & 0 & T_{13} & 0 & 0 & 0\\ -\frac{1}{2H_1} & -\frac{1}{2H_1} & -\frac{D_1}{2H_1} & \frac{1}{2H_1} & \frac{1}{2H_1} & 0\\ 0 & 0 & -\frac{1}{T_{ch_{1,1}}S_{1,1}} & -\frac{1}{T_{ch_{1,1}}} & 0 & \frac{\phi_{1,1}}{T_{ch_{1,1}}}\\ 0 & 0 & -\frac{1}{T_{ch_{1,2}}S_{1,2}} & 0 & -\frac{1}{T_{ch_{1,2}}} & \frac{\phi_{1,2}}{T_{ch_{1,2}}}\\ -K_{I_1} & -K_{I_1} & -K_{I_1}B_1 & 0 & 0 & 0\end{bmatrix}.$$

As the study of this chapter has assumed a measurement model with high re-

dundancy, the matrix C_i for Area 1 becomes

$$C_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}^\top.$$

In Area 1, the vulnerable measurements to cyber attacks are the ones of tie-line power flows $\Delta P_{tie_{1,2}}$, $\Delta P_{tie_{1,3}}$ and ΔP_{tie_1}. Thus the AGC signal ΔP_{agc_1} would be corrupted into

$$\Delta \dot{P}_{agc_1} = -k_1\left(B_1 \Delta\omega_1 + \Delta P_{tie_{1,2}} + f_{tie_{1,2}} + \Delta P_{tie_{1,3}} + f_{tie_{1,3}}\right).$$

Then the parameters regarding multivariate attacks are

$$f_1 = \begin{bmatrix} f_{tie_{1,2}} & f_{tie_{1,3}} & f_{tie_1} \end{bmatrix}^\top,$$

$$D_{1,f} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}^\top, \quad B_{1,f} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -k_1 \\ 0 & 0 & 0 & 0 & 0 & -k_1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^\top.$$

5

Robust Detection: A Novel Data-Assisted Model-based Approach

A diagnosis filter from Chapter 4 could suffice for a power system described in the linear mathematical model. However, for a more detailed description of a complex power system, complicated simulators are always exploited to predict its behavior, and mismatches do exist when we compare the mathematical model and the simulation model. In most of the literature, the diagnosis tool from such a model-based method only works effectively for a linearized model, or any mathematical model where the nonlinearity can be fully described. In this chapter the thesis considers a more realistic setting that the power system is simulated in a high-fidelity simulator in which this study aims to implement the developed diagnosis filter for detecting false data injection attacks. It would go beyond the pure model-based or data-driven viewpoints to propose a robustification scheme for the diagnosis filter with the assistance of simulation data to extract the model mismatch signatures. Till this end, an optimization-based framework is developed to (i) detect the univariate attack with non-zero transient or non-zero steady-state residual, and even track the attack values

in the steady-state behavior and isolate each intrusion; (ii) detect all the admissible multivariate attacks introduced in Definition 4.2.5 with non-zero transient or non-zero steady-state residual, by extending the theoretical results in Section 4.4.2 and Section 4.4.3 of Chapter 4. Moreover, the robustified diagnosis filter is implemented in DIgSILENT PowerFactory to validate the effectiveness of the proposed robustification scheme in detecting false data injection (FDI) attacks against the Automatic Generation Control (AGC) measurements in the three-area IEEE 39-bus system.

5.1. INTRODUCTION

5.1.1. MOTIVATIONS AND AN OVERVIEW

The principle of attack detection in intelligent power grids is to generate an output of diagnostic signal (e.g., residual) with all the inputs of available data (e.g., measurements and control signals) while decoupled from other unknown disturbances. In general, the attack detection methods can be classified into two categories: (i) the approaches that utilize the explicit mathematical models of system dynamics (referred to *model-based* methods in this chapter); (ii) the *data-driven* approaches that try to automatically learn the system characteristics from available data [33, 34]. The work of Chapter 4 has developed a diagnosis filter to detect the class of stealthy multivariate attacks. To this end, a dynamic residual generator approach has be introduced and reformulated as robust optimization programs for both the transient and steady-state behavior of the residual generator. This method is, indeed, model-based that the multivariate intrusions on the dynamics of system trajectories are captured with an explicit mathematical model representation. The results in Section 4.5 have proven its effectiveness. Now here comes another question:

Can the effectiveness of the designed model-based diagnosis filter be still ensured in a more realistic setting, or in an implementation for a real or simulated power system?

Literature on model-based and data-driven attack detection This chapter aims to address this question which may apply to all model-based attack detection methods. Before that, a brief overview on the *model-based* and *data-driven* approaches for attack detection is provided. Indeed both types have their own advantages and drawbacks [135]. The model-based methods requires that the system dynamics must be well understood. The main task of this approach is to generate a residual signal for the difference between the measured variables (output signals) and their estimates. *Observer-based* residual generator has been a major subclass of model-based detection schemes; see the review in Section 4.1.1 and the discussion in Section 4.5.3. More work can be found in [136–138] specially for detection of false data injection (FDI) attacks in intelligent power grids. Parity space [139] and parameter estimation [140] model-based methods have also been extensively investigated mainly on the topic of fault detection and isolation. For instance, extended

Kalman filter algorithms can be used to perform such an estimation for anomaly detection, identification and recovery [141]. The diagnosis filter in Chapter 4 provides a good alternative with a scalable and robust design to reveal all the admissible multivariate intrusions, while the challenge still remains as the power system models are mostly complex, high-dimensional and nonlinear. The work in [120] proposed an optimization-based filter for detecting a single fault in the scenario where the nonlinearity in the control system model can be fully described. However, developing accurate mathematical models taking all the nonlinearities and modeling errors or uncertainties into account becomes difficult or even infeasible, especially in the case that some uncertainties can not be well quantified [135].

Another technique for attack detection comes from *data-driven* approaches which do not require an explicit mathematical model of the system. Developments such as sensing technology, Internet-of-Things (IoT) and Artificial Intelligence (AI) have contributed to a more data-driven power system. Anomaly detection is mainly considered as a classification problem and there are supervised or unsupervised learning approaches for that purpose. Among all the supervised classifications, deep neural networks (DNN) [142, 143], bayesian networks [144] and Kernel machines [33] are the popular methods. For unsupervised classifications, one can find principle component analysis (PCA) and its extensions [145, 146], autoencoders [147], etc. In addition to supervised or unsupervised approaches, recent work in [148] have proposed reinforcement learning based algorithm for online anomaly detection without a prior knowledge of the attack types or models. Overall, data-driven methods are suitable for the complex and large-scale systems. However, their performance highly depends on the quantity and quality of the accessible data [135], and thus can be intractable. Besides, the required pre-processing stage (e.g., data training) may have high computational cost.

The study of this chapter is motivated to improve the diagnosis filter developed in Chapter 4 and make a step forward to an implementation in a real or simulated power system. The challenge mainly comes from the *plant-model mismatch* as discussed in Chapter 1; see Figure 1.3. A linearized mathematical model - Automatic Generation Control (AGC) for frequency dynamics has been used in Chapter 4 for the validation of the diagnosis filter. However, the real AGC model of a power system is in fact nonlinear. High-fidelity simulators (e.g., DIgSILENT PowerFactory) are

always used to describe the detailed system model to provide greater insights into dynamic behavior of the system. The mismatches still exist between the mathematical model and the simulation model, and a direct implementation of the designed diagnosis filter may fail in face such model mismatches.

5.1.2. CONTRIBUTIONS AND OUTLINE

The main objective here is to propose a robustification scheme for the model-based diagnosis filter which may encounter model mismatches when implemented in a real or simulated power system. In this regard, this study would first describe both the mathematical model and the simulation model in the instance of AGC system. Detailed simulation models have been developed in the simulator - DIgSILENT PowerFactory. Considering an augmented closed-loop system model, the model mismatches can be characterized in the framework of DAE. Finally an optimization-based approach is proposed to robustify the diagnosis filter with respect to model mismatches, with the assistance of simulation data to extract the mismatch signatures. The main contributions of this chapter are listed as follows:

(i) Firstly, the work of this chapter departs from the pure model-based or data-driven viewpoints of attack detection to further robustify the diagnosis filter to *model mismatches* with the assistance of simulation data, which would also bridge the gaps between these two types of methods. A square of $\mathcal{L}_2$-inner product with corresponding norm is introduced for the discrete-time signal in a finite time horizon, characterizing the effects of model mismatches on the filter residual output (Definition 5.3.1). Afterwards, the design of the robustification scheme is formulated as a quadratic optimization problem where the effects of model mismatches on the residual output are minimized (Definition 5.3.2 and the program (5.15) in Remark 5.3.4).

(ii) Both of the desired features of non-zero transient and non-zero steady-state behavior of the robustified diagnosis filter under univariate or multivariate attacks are investigated. In particular, as identified by Lemma 5.3.5, a diagnosis filter with the linear constraints from (5.16) can have non-zero steady-state residual and also track the univariate attack value (if there is no model mismatch, it recovers the exact attack value). For the case of sufficient compu-

tational resources, an approach is also provided to isolate each intrusion in a multivariate attack by utilizing a bank of diagnosis filters (Remark 5.3.6). Extensions of the theoretical results from Chapter 4 are also made to robustify the diagnosis filter to model mismatches while keeping sensitive to any plausible disruptive multivariate attacks (Corollary 5.3.7 and Corollary 5.3.8).

The process of diagnosis filter construction and validation is concluded in Algorithm 2. Besides, the effectiveness of the proposed robust scheme is validated on the three-area IEEE 39-bus system. Numerical results illustrate that the robustified diagnosis filter implemented in PowerFactory can successfully generate alerts for both univariate and multivariate attacks, while a filter without robustification to model mismatches may fail by triggering "false alarms".

Section 5.2 presents both the mathematical model in DAE and the simulation model in PowerFactory. Model mismatches are highlighted and quantified. Section 5.3 proposes the robust scheme for the diagnosis filter design where an optimization framework is introduced for univariate and multivariate attacks, after characterizing the model mismatch signatures. Numerical results for validation are reported in Section 5.4.

5.2. SYSTEM MODELING

In a power system, the physical plant is operated by a digital controller that receives measurements as inputs and sends control actions to the actuators through communication networks (e.g., SCADA). As shown in Figure 4.1 of Section 4.5.1, the typical structure of a closed-loop system has the four main components: (i) the physical plant, (ii) the communication network, (iii) the digital feedback controller, and (iv) an anomaly detector (diagnosis filter) [16]. In this section, both the mathematical model and the simulation model in a simulator are described that the first three components (i)-(iii) are mainly considered. In Section 5.3 the improved diagnosis filter with a robustification to possible model mismatches will be proposed.

5.2.1. MATHEMATICAL MODEL DESCRIPTION

The work of Chapter 4 has introduced the linear state-space representations (4.3) (4.4) for the closed-loop system. Besides, as noted by Remark 4.2.1, transferring

from (4.3) to (4.4) is without loss of generality that it is rich enough to subsume a dynamic controller architecture as well. Here such details are illustrated. Consider the general system model in Figure 4.1. Suppose the control signal is implemented as a feedback controller described by the discrete-time dynamics

$$\begin{cases} X_c[k+1] = A_c X_c[k] + B_c Y[k], \\ u[k] = C_c X_c[k] + D_c Y[k], \end{cases} \tag{5.1}$$

where the input is the dynamical system measurements $Y[k] \in \mathbb{R}^{n_Y}$ at time step $k \in \mathbb{N}$, the output is the control signal $u[k] \in \mathbb{R}^{n_u}$ and the internal state of the controller is denoted by $X_c \in \mathbb{R}^{n_c}$. When a false data injection attack corrupts the measurements, it affects the dynamics of the controller and consequently the involved physical system. To study the controller dynamics together with the original dynamical system, one can augment the states of the system (4.3) together with the controller's as $\hat{X} := \left[X^\top \; X_c^\top\right]^\top$. An augmented output signal can be also presented as $\hat{Y} = [Y^\top \; u^\top]^\top$, which is available to the power system operators. Thus the dynamics of the closed-loop system can be described by

$$\begin{cases} \hat{X}[k+1] = \hat{A}_{cl}\hat{X}[k] + \hat{B}_d d[k] + \hat{B}_f f[k], \\ \hat{Y}[k] = \hat{C}\hat{X}[k] + \hat{D}_f f[k]. \end{cases} \tag{5.2}$$

where the involved matrices are defined as

$$\hat{A}_{cl} := \begin{bmatrix} A_x + B_u D_c C & B_u C_c \\ B_c C & A_c \end{bmatrix}, \quad \hat{B}_d := \begin{bmatrix} B_d \\ 0 \end{bmatrix}, \quad \hat{B}_f := \begin{bmatrix} B_u D_c D_f \\ B_c D_f \end{bmatrix},$$

$$\hat{C} := \begin{bmatrix} C & 0 \\ D_c C & C_c \end{bmatrix}, \quad \hat{D}_f := \begin{bmatrix} D_f \\ D_c D_f \end{bmatrix}.$$

In this view, the augmented system (5.2) shares the same architecture as (4.4) studied in Chapter 4 for the instance of static feedback controller. Following the process of transferring (4.4) to (4.5), it can be shown that the state-space description (5.2) is still a particular case of the DAE model. By introducing a time-shift operator $q : q\hat{X}[k] \to \hat{X}[k+1]$, one can fit (5.2) into the formulation of (4.5), namely,

$$H(q)x[k] + L(q)y[k] + F(q)f[k] = 0, \tag{5.3}$$

where $x := [\hat{X}^\top \; d^\top]^\top$ represents the unknown signals of closed-loop system states and disturbances, $y := \hat{Y}$ contains all the available data $Y[\cdot]$ and $u[\cdot]$ for the operator and also the diagnosis filter as Figure 4.1 indicates. Similarly, let n_x and n_y be

the dimensions of $x[\cdot]$, $y[\cdot]$ and n_r be the number of rows in (5.3). Then H, L, F are polynomial matrices in terms of the time-shift operator q with n_r rows and n_x, n_y, n_f columns separately, by defining,

$$H(q) := \begin{bmatrix} -qI + \hat{A}_{cl} & \hat{B}_d \\ \hat{C} & 0 \end{bmatrix}, \quad L(q) := \begin{bmatrix} 0 \\ -I_{(n^Y+n^u)} \end{bmatrix}, \quad F(q) := \begin{bmatrix} \hat{B}_f \\ \hat{D}_f \end{bmatrix}.$$

Remark 5.2.1 (Communication network effects)**.** *The communication network is responsible for the timely data delivery between the digital controller and the sensors or actuators in the power system. Network features such as the communication delay and the packet loss intrinsically affect the behavior of the physical system, as discussed in Chapter 3. In the mathematical models above, these effects are not considered. Instead, this study focuses on attacks due to malicious adversaries and assumes a reliable communication network for the ease of the diagnosis filter design and validation. However, the closed-loop framework* (5.2) *is still rich enough to include some of these features. For instance, an augmented discrete-time model can be formulated taking into account the deterministic or random communication delays; see [149] for such an analysis.*

5

THE MODELING INSTANCE: AUTOMATIC GENERATION CONTROL

Recall the modeling instance of power system dynamics: Automatic Generation Control (AGC) system under multivariate attacks in Section 4.3. The AGC is an automatic closed-loop system that regulates power system frequency by tuning the setpoints of the participating generators. For a distributed multi-area power system in Figure 4.2, the AGC block in each area collects the frequency and tie-line power flow measurements and sends back control signals to the participating generators, through SCADA networks mostly with DNP 3.0 protocol. After receiving measurements, the control center in Area i calculates an area control error (ACE) signal,

$$ACE_i = \beta_i(\omega_i - \omega_0) + (P_{tie_i} - P_{tie_0}), \tag{5.4}$$

where β_i is the frequency bias, ω_i and P_{tie_i} denote the frequency and tie-line power flow measurements of Area i, and ω_0 and P_{tie_0} correspond to the nominal or scheduled values. The ACE value defines the power to compensate and the frequency to restore in the event of load - generation imbalance. With the input of ACE_i, the AGC controller generates an output control signal for the participating generators

in that area to track the load changes. This is usually a integral action which can be expressed as

$$\Delta \dot{P}_{agc_i} = K_{I_i} ACE_i, \tag{5.5}$$

where K_{I_i} is the integral coefficient for the AGC block in Area i, and ΔP_{agc_i} represents the AGC output signal that is feeding into the governors of the generators in Area i according to each generator's participating factor ($\phi_{i,g}$ in (4.8b)). In the work of AGC analysis, usually, each area of a power grid is represented by a linearized model comprised of equivalent rotating mass, governors and turbines, and is decoupled from the voltage dynamics (automatic voltage regulator loop) [118]. With the linearized model together with (5.4) and (5.5), the mathematical model of AGC has been derived in the form of (4.4) (and also (5.2)) after discretization.

To be noted, till now the linear AGC system descriptions are used, as the study restricts the attention on the design of scalable diagnosis filter for detecting all admissible multivariate attacks. However, the models of complex power systems are mostly nonlinear. In fact, though the linearized AGC model used for diagnosis filter test and validation is sufficiently accurate for studies, it is still an abstract model which lacks some essential details compared to a full and detailed system model. Nonlinear versions of AGC can be found in the work of [112, 150, 151]. In the following subsection, the simulator DIgSILENT PowerFactory is used to build a more detailed simulation model of an IEEE benchmark system equipped with AGCs.

5.2.2. SIMULATION MODEL IN DIGSILENT POWERFACTORY

To demonstrate the AGC operation, the simulation models in PowerFacotry have been built for the IEEE 39-bus system (Figure 4.2 of Chapter 4). Figure 5.1 shows the test system model equipped with AGC in PowerFactory. This power system consists of 39 buses, 12 two-winding power transformers, 34 lines and 19 loads, 10 generators where 7 of them are participating AGC as depicted in Figure 5.1, representing a three-area transmission network.

In the simulations, the dynamic generator model consists of a synchronous machine, along with automatic voltage regulator (AVR) as excitation system in the type of IEEE Type 1, turbine-governor model (GOV) in the type of IEEE Type G1 (steam turbine) or IEEE Type G3 (hydro turbine). These controllers are part of all the participated generators in the test case. AVR helps with regulating the voltage

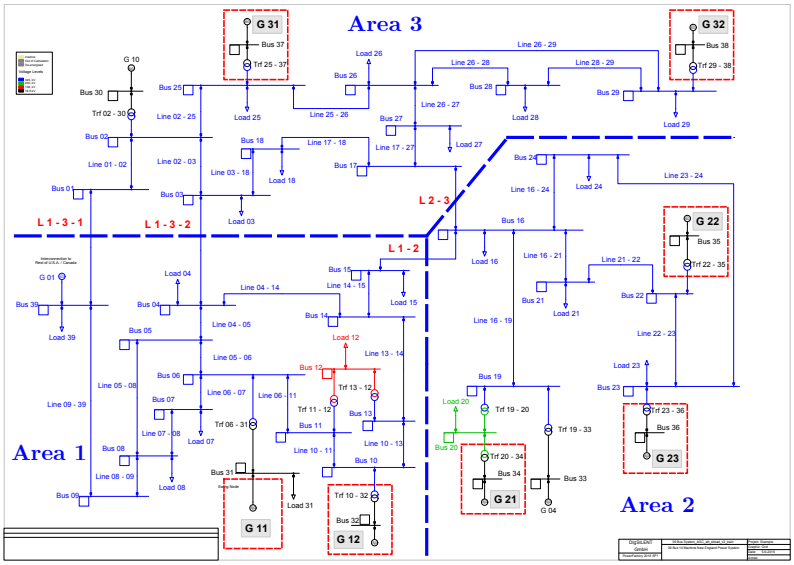


Figure 5.1: Three-area 39-bus system equipped with AGC in DIgSILENT PowerFactory.

of the system by changing field winding voltage. The governor is used to regulate initial frequency variations by a type of droop control.

According to Figure 5.1, the 39-bus transmission network is divided into three areas: Area 1 consists of 2 generators (G 11, G 12) participating in AGC and 7 loads; Area 2 has 3 generators (G 21, G 22, G 23) for AGC and 5 loads; Area 3 contains 2 generators (G 31, G 32) participating in AGC and 7 loads. Transmission lines called tie-lines (L 1-3-1, L 1-3-2, L 1-2, L 2-3) connect areas. Each area has its own AGC controller to regulate the frequency of each area and the exchanged tie-line power flows. Each AGC in that area collects the measurements of frequency and tie-line power flows exported from that area. The AGC controller then uses these measurements to calculate the ACE signal in (5.4).

The AGC controllers have been developed by the *DIgSILENT Simulation Language (DSL)*. Figure 5.2 shows the *composite frame* of AGC for Area 1 in PowerFactory. This frame show the connections between the inputs and outputs of AGC: in Area 1, the first four slots in the left are for measuring the frequency at Bus 04 of Area 1 and the tie-line power flows in L 1-3-1, L 1-3-2 and L 1-2 at the sides of Area 1; then these measurements are inputs for the AGC (the slot in the right). Using the PowerFactory libraries, the AGC *block definitions* for all areas can be created. For instance,

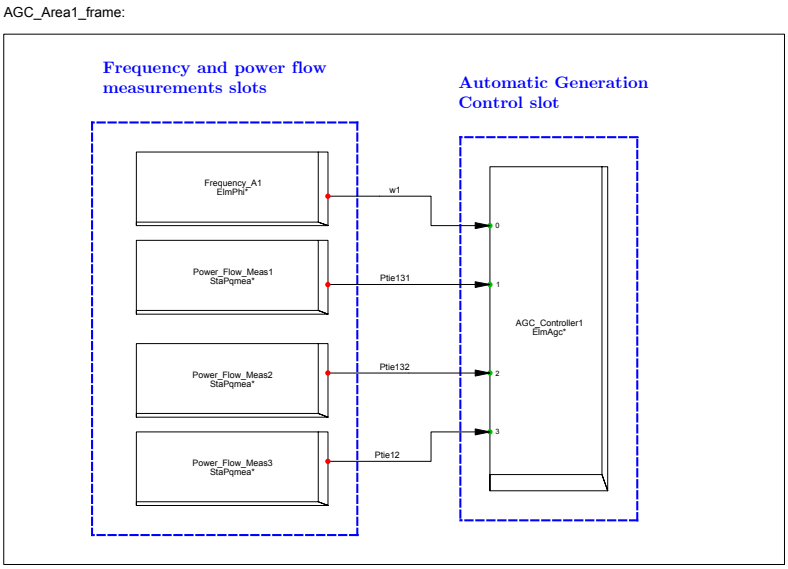


Figure 5.2: The composite frame of AGC for Area 1 in DIgSILENT PowerFactory.

Figure 5.3 illustrates the block definition of AGC in Area 1. This block definition has four sub-blocks,

- *frequency deviations* block where the frequency deviations in p.u. multiplied by the frequency bias in ACE of Area 1 are calculated;
- *tie-line power flows deviations* block which similarly computes the tie-line power flow deviations on the side of Area 1 for the power part in ACE of (5.4) after a normalization in p.u.;
- *AGC controller* block which performs the calculation in (5.5) to generate the control signal ΔP_{agc_1}. To be noted, due to saturation, the limits of the minimum ΔP_{agc}^{min} and maximum ΔP_{agc}^{max} are added for the control signal,

$$\Delta P_{m,agc_i} = \begin{cases} \Delta P_{agc}^{min} & \text{if } \Delta P_{agc_i} \leq \Delta P_{agc}^{min}, \\ \Delta P_{agc_i} & \text{if } \Delta P_{agc}^{min} < \Delta P_{agc_i} < \Delta P_{agc}^{max}, \\ \Delta P_{agc}^{max} & \text{if } \Delta P_{agc_i} \geq \Delta P_{agc}^{max}, \end{cases}$$

  where $\Delta P_{m,agc_i}$ denotes the real AGC signal for generators in PowerFactory;
- *AGC output signals* block where the output signals for all the participated generators in Area 1 are calculated based on each generator's participating factor.

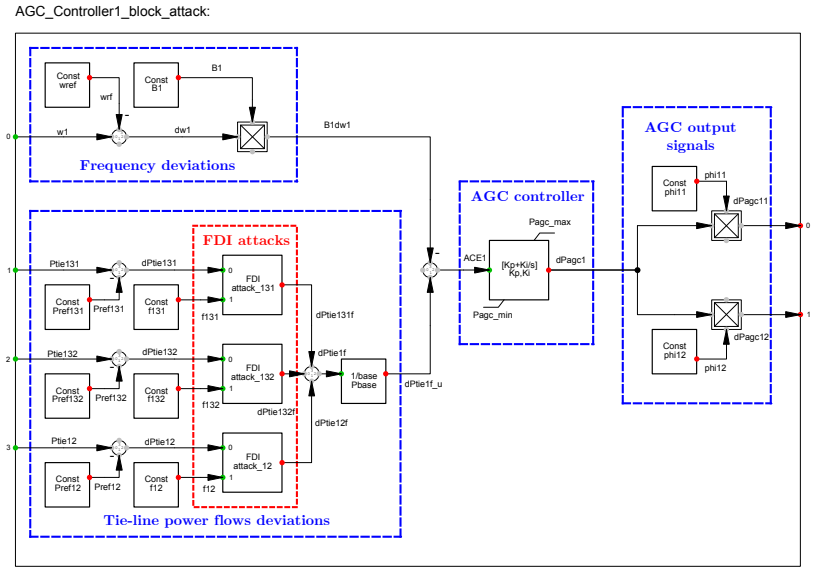


Figure 5.3: The block definition of AGC for Area 1 in DIgSILENT PowerFactory.

The above *block definitions* are modeled by using *Standard Macros* of PowerFactory global Library. Moreover, in Figure 5.3, another *block definition* (in red diagram) corresponds to the false data injection attack model for the study of multivariate attack detection in Chapter 4 and 5.

- *FDI attacks* block where the multivariate attacks in Section 4.2.2 of Chapter 4 are simulated. Each FDI attack block capture the feature of stationary attack in Assumption 4.2.4 that it can add a step input to the existing signal. One can also specify the time of the attack and the attack values. This block definition is achieved by using the *digexfun* interface. With *digexfun*, we can define our own DSL function (in C++) and create a dynamic link library *digexfun_*.dll* that PowerFactory can load.

In the real implementation, the inputs of frequency and tie-line power flow measurements for the AGC controller are delivered at specific time intervals. The calculated mechanical power setpoints (i.e., the AGC output signals) are then delivered to the participating generators in each area. Thus this process has a discrete-time nature. Besides, the data of measurements and AGC output signals are transmitted through SCADA communication networks. In this chapter, as Remark 5.2.1

states, the study is mainly focusing on the modeling of the physical power network, AGC controllers and attacks, and an ideal communication network is assumed.

5.2.3. MODEL MISMATCHES

As discussed Section 5.1.1, mismatches do exist between the abstract mathematical model and the detailed simulation model (of PowerFactory) in several aspects. This makes the implementation of the developed model-based diagnosis filter in a real or simulated power system non-trivial. From the perspective of system modeling, this study still aims to describe the *model mismatches* in the framework of DAE such that the effects of model mismatches on the diagnosis filter outputs can be described. When both of the DAE mathematical descriptions (5.2) and (5.3) are available to the diagnosis filter design, one can add an additional term to describe the detailed model of interest in the simulator as follows,

$$E(x[k]) + H(q)x[k] + L(q)y[k] + F(q)f[k] = 0\,, \tag{5.6}$$

where $E(x[k])$ represents the model mismatch between these two models and it is an "unknown" signal with n_r dimensions and also it is function of $x[k]$ which consists of internal states and natural disturbances. Note the difference between (5.3) and (5.6). Intuitively the model mismatch term $E(x[k])$ would affect the performance of the diagnosis filter which fully depends on (5.3), if it is implemented for (5.6) directly. In what follows it is shown how to use the available knowledge of the mathematical model (the polynomial matrices $H(q)$, $L(q)$ and $L(q)$) and the simulation data which can extract the signature of $E(x[k])$ to robustify the diagnosis filter to possible model mismatches.

5.3. A NOVEL DATA-ASSISTED MODEL-BASED DETECTION APPROACH

5.3.1. PRELIMINARIES FOR ROBUST ATTACK DETECTION

The diagnosis filter is still restricted to the type of dynamic residual generator in the form of linear transfer function $r_D[k] := R(q)y[k]$ where r_D is the residual signal of the diagnosis filter with the inputs $y[\cdot]$ and $R(q)$ is a transfer operator that has a formulation of $R(q) := a(q)^{-1}N(q)L(q)$; recall Section 4.4.1. $N(q)$ with the dimen-

sion of n_r and a predefined order d_N is the design variable for the diagnosis filter construction, if the denominator $a(q)$ with sufficient order is determined. By multiplying $a(q)^{-1}N(q)$ in the left of (5.6), we can further obtain

$$\begin{aligned} r[k] &= a(q)^{-1}N(q)L(q)y[k] \\ &= -\underbrace{a(q)^{-1}N(q)H(q)x[k]}_{\text{(I)}} - \underbrace{a(q)^{-1}N(q)F(q)f[k]}_{\text{(II)}} - \underbrace{a(q)^{-1}N(q)E(x[k])}_{\text{(III)}}, \end{aligned} \tag{5.7}$$

where term (I) in (5.7) is the part introduced by $x[\cdot]$ of system states and disturbances. Term (II) is the desired contribution from the attack $f[\cdot]$. Different from (4.14), in (5.7) there is term (III) that denotes the effects of model mismatches on the residual output. This can be characterized by the following definition.

5

Definition 5.3.1 (Effects of model mismatches on filter residual)**.** *For the description of effects from the model mismatches on the residual generator output, let us define*

$$e_x[k] := E(x[k]), \quad r_e[k] := -a(q)^{-1}N(q)e_x[k], \quad \forall k \in \mathbb{N}, \tag{5.8}$$

The signal $e_x[\cdot]$ is called the model mismatch signature*. Next, let us denote the space of a discrete-time signal with n dimensions over a period of T (i.e., $k \in \{1, \cdots, T\}$) to $\mathbb{R}^{n\times T}$ by $\mathcal{M}_T^n$. Thus for the signal $r_e[\cdot] \in \mathcal{M}_T^1$, a square of $\mathcal{L}_2$-inner product with corresponding norm is introduced such that*

$$\|r_e\|_{\mathcal{L}_2}^2 := \langle r_e,\, r_e\rangle, \quad \langle e,\, g\rangle := \sum_{k=1}^{T} e^\top[k]g[k], \quad \forall e,\, g \in \mathcal{M}_T^n. \tag{5.9}$$

This chapter aims to design a class of residual generator which still keeps sensitive to all disruptive attacks and also ensure the effects from model mismatches on the residual output being minimized. For minimizing the effects of model mismatches, from Definition 5.3.1, the goal now becomes to minimize $\|r_e\|_{\mathcal{L}_2}^2$ in an optimization-based framework. The above diagnosis filter design can be characterized by a class of residual generator which has the following features.

Definition 5.3.2 (Residual generator robust to model mismatches)**.** *Consider a linear residual generator represented via a polynomial vector $N(q)$ for a given $a(q)$. This residual generator is robust with respect to model mismatches introduced in* (5.6) *and*

can detect all admissible disruptive attacks in Definition 4.2.5, if

$$\begin{cases} (I) & N(q)H(q) = 0, \\ (II) & N(q)F(q)f \neq 0, \quad \forall f \in \mathcal{F}, \\ (III) & \|r_e\|^2_{\mathcal{L}_2} \to 0, \quad \forall T \in \mathbb{N}, \end{cases} \tag{5.10}$$

where $\mathcal{F}$ is the disruptive attack set similar to the one in Definition 4.2.5. Term (III) aims to find an appropriate polynomial vector $N(q)$ to minimize the effects from the model mismatch on the residual generator output.

The first two polynomial equations in (5.10) of Definition 5.3.2 can be characterized as a feasibility problem of a finite robust program. Recall Lemma 4.4.2 and the robust program reformulations in Section 4.4.2 and 4.4.3 for both the transient and steady-state behavior of the residual generator under multivariate attacks. In the following of this section, the robust programs are extended to handle term (III).

5

5.3.2. DIAGNOSIS FILTER FOR A UNIVARIATE ATTACK

To explain the process of diagnosis filter construction, let us first consider a univariate attack scenario where only one measurement gets attacked ($n_f = 1$), while in the next subsection the case would be extended to multivariate attacks ($n_f > 1$). In what follows it would be shown how the simulation data can be used to extract the model mismatch signatures to assist in building a diagnosis filter robustified to possible model mismatches.

Definition 5.3.3 (Model mismatch signature extraction)**.** *Recall the DAE framework (5.3) (5.6) where $x = [\hat{X}^\top \; d^\top]^\top$ contains unknown signals of system states and natural disturbances. Let $x_i[\cdot]$ denote one instance of $x[\cdot]$. Thus for each $x_i[\cdot]$, from the data of DIgSILENT PowerFactory's simulation and the DAE's computation, one can have a specific model mismatch signature $e_{x_i}[\cdot]$. For $e_{x_i}[\cdot] \in \mathcal{M}_T^{n_r}$, using Definition 5.3.1, a* mismatch signature matrix *$e_{x_i} \in \mathbb{R}^{n_r \times T}$ is introduced such that*

$$e_{x_i} = \begin{bmatrix} e_{x_i}[1] & e_{x_i}[2] & \cdots & e_{x_i}[T] \end{bmatrix}. \tag{5.11}$$

Recall that the linear operator q acts as a time-shift operator: $qe_{x_i}[k] \to e_{x_i}[k+1]$. Thus it can be approximately translated as a matrix left-shift operator for matrix

e_{x_i}: $qe_{x_i} \to e_{x_i} D$ where D is a square matrix of order T,

$$D = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & \ddots & \vdots & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{5.12}$$

Using the matrices (5.11) (5.12) above, it can be further obtained that

$$N(q)e_{x_i} = \sum_{j=0}^{d_N} N_j q^j e_{x_i} = \bar{N} \begin{bmatrix} I_{n^r} \\ qI_{n^r} \\ \vdots \\ q^{d_N} I_{n^r} \end{bmatrix} e_{x_i} = \bar{N} D_{x_i}, \quad D_{x_i} = \begin{bmatrix} e_{x_i} \\ e_{x_i} D \\ \vdots \\ e_{x_i} D^{d_N} \end{bmatrix}, \tag{5.13}$$

where $\bar{N} = \begin{bmatrix} N_0 & N_1 & \cdots & N_{d_N} \end{bmatrix}$ as defined in Lemma 4.4.2. Thus according to the notation of $\mathcal{L}_2$-inner product, for one instance of the unknown signal $x_i[\cdot]$, further it will arrive at

$$\| r_{e_{x_i}} \|_{\mathcal{L}_2}^2 = \bar{N} Q_{x_i} \bar{N}^\top, \quad Q_{x_i} = D_{x_i} G {D_{x_i}}^\top, \tag{5.14}$$

where G is a square matrix of order T that $G(i,j) = \left\langle a(q)^{-1} b_i[\cdot],\ a(q)^{-1} b_j[\cdot] \right\rangle$. Note $b_i[\cdot]$ is a discrete-time "basic" signal where the only non-zero value (= 1) occurs at the time instance $i \in \{1, \cdots, T\}$, i.e., $b_i[i] = 1$ and $b_i[k] = 0$ for $k \neq i$.

Remark 5.3.4 (Robust diagnosis filter to univariate attack)**.** *In order to robustify the diagnosis filter, the diagnosis filter can be* trained *under more than one unknown signals $x_i[\cdot]$, i.e., $\{x_i[\cdot]\}_{i=1}^n$. Note that for each $x_i[\cdot]$, if there is no attack, the system trajectories of dynamics only depend on the input of natural disturbance in the closed-loop system, say $d_i[\cdot]$. Thus for each disturbance signature $d_i[\cdot]$ (and correspondingly $x_i[\cdot]$) when there is no attack, the model mismatch signature $e_{x_i}[\cdot]$ and also the matrices e_{x_i}, D_{x_i}, Q_{x_i} can be computed from (5.11) to (5.14). Next, in light of (5.10), the robust diagnosis filter design can be formulated as an optimization program where the objective can be minimizing $\bar{N}(\frac{1}{n}\sum_{i=1}^n Q_{x_i})\bar{N}^\top$ (average-cost viewpoint) or $\max_{i \le n}(\bar{N} Q_{x_i} \bar{N}^\top)$ (worst-case viewpoint). Consider the polynomial matrices $H(q) = \sum_{i=0}^1 H_i q^i$ and $F(q) = F$, where $H_i \in \mathbb{R}^{n_r \times n_x}$ and $F \in \mathbb{R}^{n_r \times n_f}$ are constant matrices. This study takes the former "average performance" objective and proposes the*

following optimization program to robustify the diagnosis filter by minimizing the effects from model mismatches on residual output,

$$\begin{aligned} \min_{\bar{N}} \quad & \bar{N}\Big(\frac{1}{n}\sum_{i=1}^{n} Q_{x_i}\Big)\bar{N}^\top \\ s.t. \quad & \bar{N}\bar{H} = 0, \\ & \big\|\bar{N}\bar{F}_f\big\|_\infty > \nu. \end{aligned} \tag{5.15}$$

where $\|\cdot\|_\infty$ *denotes the infinite vector norm in the constraint, and*

$$\bar{H} := \begin{bmatrix} H_0 & H_1 & 0 & \cdots & 0 \\ 0 & H_0 & H_1 & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & H_0 & H_1 \end{bmatrix}, \quad \bar{F}_f := \begin{bmatrix} F & 0 & \cdots & 0 \\ 0 & F & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & F \end{bmatrix}.$$

In (5.15), the two constraints are desired features for non-zero transient behavior of the diagnosis filter under a univariate attack. These two characterizations are similar to the ones in Lemma 4.4.2 for multivariate attacks. To be precise, the proposed optimization (5.15) is not a quadratic program (QP) due to the last non-convex constraint. However, as explained by a similar argument in [120, Lemma 4.3], one can view (5.15) as a family of $d_N + 1$ standard QPs.

Chapter 4 has considered both the transient and steady-state behavior of the diagnosis filter. The following lemma studies the steady-state behavior of the diagnosis filter for univariate attacks.

Lemma 5.3.5 (Steady-state behavior and univariate attack tracking)**.** *Consider a univariate attack being stationary as Assumption 4.2.4, a diagnosis filter from the optimization program* (5.15) *but with the following linear program characterizations for the two constraints can have non-zero steady-state residual output that approximates the attack value* f *when there exist model mismatches (while if* $e_x[\cdot] \equiv 0$*, it recovers the exact attack value* f *instead of approximation),*

$$\begin{cases} \bar{N}\bar{H} = 0, \\ -a(1)^{-1}\sum_{i=0}^{d_N} N_i F = 1, \end{cases} \tag{5.16}$$

Proof. Recall that $N(q)H(q) = \bar{N}\bar{H}[I_{n^x},\ qI_{n^x},\ \cdots,\ q^{d_N+1}I_{n^x}]^\top$. Thus if $\bar{N}\bar{H} = 0$, the diagnosis filter becomes $r_D[k] = -a(q)^{-1}N(q)f[k] - a(q)^{-1}N(q)e_x[k]$. For the case

that there is no model mismatch $e_x[\cdot] \equiv 0$, the steady-state value of the filter residual under the univariate attack would be $-a(q)^{-1}N(q)F(q)f|_{q=1}$. Note that $N(1)F(1) = \sum_{i=0}^{d_N} N_i F$. Thus with (5.16) the residual output recovers the univariate attack value in the steady state. When $e_x[\cdot] \neq 0$, using the optimization program (5.15) but replacing the constraints with these linear two in (5.16), the output of the residual during the steady-state behavior could approximate the attack value f. ■

In this subsection the univariate attack scenario is considered. For multivariate attacks where $n_f > 1$, Chapter 4 has designed a diagnosis filter with a synthesized residual output for the whole closed-loop system. In fact, if the computational resources are sufficient, one can also build a bank of diagnosis filters formulated in (5.15) where each filter aims to detect one particular univariate attack.

Remark 5.3.6 (Attack isolation)**.** *The diagnosis filter from* (5.15) *is designed for one univariate attack. For multivariate intrusions, an alternative is to build a bank of diagnosis filters where each diagnosis filter is associated with one of the intrusions and decoupled from others, by considering the DAE,*

$$E(x[k]) + \begin{bmatrix} H(q) & F_{-i}(q) \end{bmatrix} \begin{bmatrix} x[k] \\ f_{-i}[k] \end{bmatrix} + L(q)y[k] + F_i(q)f_i[k] = 0, \tag{5.17}$$

where $F_{-i}(q)$ is the polynomial matrix that includes all columns of $F(q)$ except the i-th one, and similarly $f_{-i}[k]$ contains all the elements of $f[k]$ except the i-th one. Then the i-th diagnosis filter can be designed using the same approach as Remark 5.3.4 for the i-th intrusion while isolating the effects from others. The i-th intrusion can be identified by the i-th diagnosis filter since the other diagnosis filters keeps insensitive to this intrusion. Besides, with (5.16) *in Lemma 5.3.5, it can track the i-th attack's value in steady-state behavior. Similar lines of such arguments can be found in [120, Remark 4.1] and [118, Section V.B].*

5.3.3. DIAGNOSIS FILTER FOR MULTIVARIATE ATTACKS

This section aims to design a synthesized diagnosis filter for the multivariate attacks while there exist model mismatches, extending the work in Section 4.4 of Chapter 4 towards an implementation in a real or simulated power system. With the process above for model mismatch signatures extraction and effects on residual output minimization, the following results are derived for desired features of (i) a non-zero

transient and (ii) a non-zero steady-state behavior of the residual generator in the presence of multivariate attacks and also some model mismatches.

Corollary 5.3.7 (Quadratic program for diagnosis filter with non-zero transient behavior under model mismatches)**.** *Recall the finite reformulation in Theorem 4.4.3 and linear program relaxation in Corollary 4.4.4. Given $i \in \{1, \ldots, 2d_N+2\}$, for each i, consider the program with quadratic objective and linear constraints*

$$\begin{aligned} \min_{\bar{N},\lambda} \quad & \bar{N}(\frac{1}{n}\sum_{i=1}^{n} Q_{x_i})\bar{N}^\top, \\ s.t. \quad & b^\top \lambda \geq \gamma_i, \\ & (-1)^i N_{\lfloor i/2 \rfloor} F F_{\mathrm{b}} = \lambda^\top A, \\ & \bar{N} \in \mathcal{N},\ \lambda \geq 0, \end{aligned} \tag{QP$_{1,i}$}$$

where $\lfloor \cdot \rfloor$ is the ceiling function that maps the argument to the least integer. The attack basis matrix F_b, the matrix A and vector b in the polytopic set $\mathcal{A}$ of multivariate attacks, and the symmetric set $\mathcal{N}$ for the design variable $\bar{N}$ are referred to Chapter 4. In particular, if for any $i \in \{1, \ldots, 2d_N+2\}$, one can find a $\gamma_i > 0$ that (QP$_{1,i}$) is still feasible, then the solution to QP$_{1,i}$ offers a robust residual generator that detects all admissible disruptive attacks introduced by Definition 4.2.5 during transient behavior and also keeps the effects from model mismatches minimized, satisfying the terms listed in Definition 5.3.2.

Note that indeed Corollary 5.3.7 is based on Corollary 4.4.4 and Remark 5.3.4. For a better illustration, Algorithm 2 concludes the diagnosis filter construction and validation process with desired non-zero transient behavior in the presence of multivariate attacks and model mismatches, according to Corollary 5.3.7. With the results regarding steady-state behavior of the diagnosis filter in Section 4.4, this study can have another program for the diagnosis filter construction,

Corollary 5.3.8 (Quadratic program for diagnosis filter with non-zero steady-state behavior under model mismatches)**.** *Following Theorem 4.4.7 for exact convex re-*

Algorithm 2 Diagnosis filter construction and validation for multivariate attacks under model mismatches

1) **Pre-training**: For all $i \in \{1, \ldots, 2d_N + 2\}$, solve ($\mathrm{LP}_i$) in Corollary 4.4.4 of Section 4.4.2. Check if there exists $\gamma_i^\star > 0$ and find the maximum of $\{\gamma_0^\star, \gamma_1^\star, \cdots, \gamma_{2d_N+2}^\star\}$.

2) **Training phase**:

   (i) For each instance of disturbance d_i when there is no attack, run the PowerFactory simulations and computes DAEs in Matlab. Calculate the mismatch signature $e_{x_i}[\cdot]$ and also the matrices D_{x_i}, Q_{x_i} according to (5.11) - (5.14).

   (ii) For a number of n instances of disturbance, perform the process in (i).

   (iii) Set the initial value of γ_i in (5.15) to be $\max_{\{i\leq 2d_N+2\}} \gamma_i^\star$ from **pre-training**. Solve (5.15) with the calculated matrix Q_{x_i}. Tune the value γ_i until it reaches maximum.

3) **Testing phase:** For another instance of disturbance with the same patter as the ones in **training phase**, also add the attack scenario, run the PowerFactory simulations and run the diagnosis filter with the results of design variables from **training phase**. The inputs of the diagnosis filter are the simulation data from PowerFactory. Check the performance of the filter.

formulations, consider the program with quadratic objective and linear constraints

$$\begin{aligned} \min_{\bar{N},\lambda} \quad & \bar{N}\Big(\frac{1}{n}\sum_{i=1}^{n} Q_{x_i}\Big)\bar{N}^\top, \\ s.t. \quad & b^\top \lambda \geq \mu, \\ & \bar{N}\bar{F} = \lambda^\top A, \\ & \bar{N} \in \mathcal{N},\ \lambda \geq 0, \end{aligned} \tag{QP$_2$}$$

where attack basis matrix F_b, the matrix A and vector b in the polytopic set $\mathcal{A}$ of attacks, the symmetric set $\mathcal{N}$ for the design variable $\bar{N}$ are still referred to Chapter 4. Besides, $\bar{F} = \begin{bmatrix} FF_\mathrm{b} & FF_\mathrm{b} & \cdots & FF_\mathrm{b} \end{bmatrix}^\top$ is the matrix defined in Lemma 4.4.6. If one can

find $\mu > 0$, then the solution to QP_2 offers a robust residual generator that detects all admissible disruptive attacks introduced by Definition 4.2.5 with a non-zero steady-state residual level in a long-time horizon, and also keeps the effects from model mismatches minimized, satisfying the terms in Definition 5.3.2.

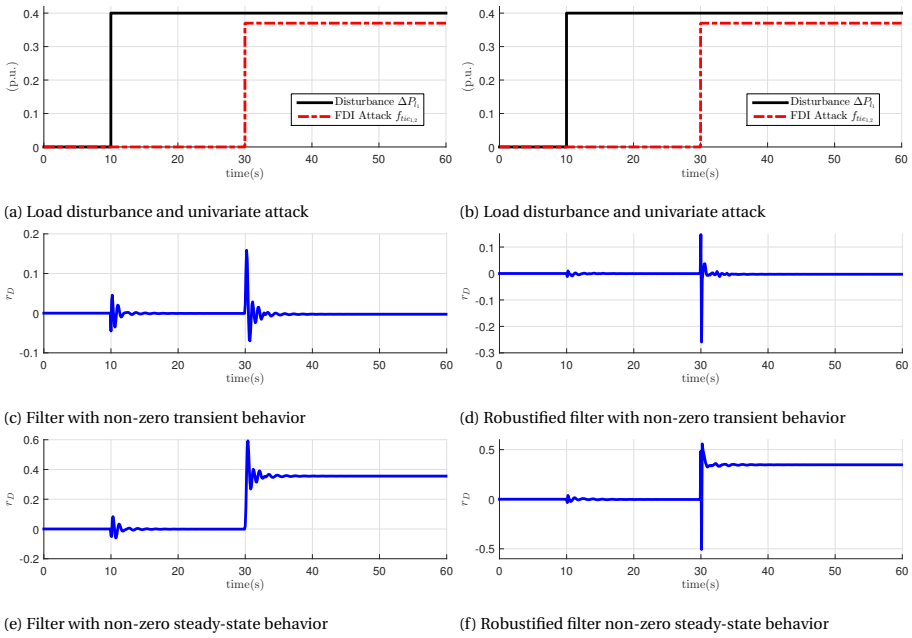
It can be seen that Corollary 5.3.8 is extended from Theorem 4.4.7 and Remark 5.3.4. The procedure of the diagnosis filter construction and validation with desired non-zero steady-state residual in the presence of multivariate attacks and model mismatches is similar to Algorithm 2. In the pre-training phase, one needs to solve (4.25a) in Theorem 4.4.7 to see if there exists $\mu^\star > 0$. If yes, from the implication of Theorem 4.4.7 and Remark 4.4.8, it can be concluded that, the diagnosis filter detects all the admissible attacks with non-zero steady-state residual output in a long-term behavior. Assisted with the model mismatch signatures extracted from the simulation data, the program (QP_2) also keeps the effects of model mismatches on the filter residual output minimized.

5.4. NUMERICAL RESULTS

5.4.1. TEST SYSTEM AND ROBUST DETECTOR DESCRIPTION

To validate the effectiveness of the proposed approach, the data-assisted diagnosis filter has been implemented to detect FDI attacks on the AGC measurements of the three-area 39-bus system in DIgSILENT PowerFactory. As depicted in Figure 5.1, all the three areas are equipped with AGC for load frequency control. Area 1, 2 and 3 contain, respectively, 2, 3 and 2 participating generators, and the generators in each area have equal participating factors. The AGC parameters for each area are referred to [127], and the specifications of the three-area system model in PowerFactory are available at [152] with the base of 100 MVA and 60 Hz. The linear mathematical model of frequency dynamics has been developed in Section 4.3, while the simulation model of the three-area system with AGC in PowerFactory has been detailed in Section 5.2.2.

In the simulations the robustified diagnosis filter is used to detect both univariate and multivariate attacks following Algorithm 2 in Section 5.3. To obtain the matrix Q_x, one needs to run the simulations of DAE and PowerFactory with the same input $d[\cdot] := \Delta P_l[\cdot]$ under normal scenarios (no attacks), where ΔP_l denotes the



(a) Load disturbance and univariate attack

(b) Load disturbance and univariate attack

(c) Filter with non-zero transient behavior

(d) Robustified filter with non-zero transient behavior

(e) Filter with non-zero steady-state behavior

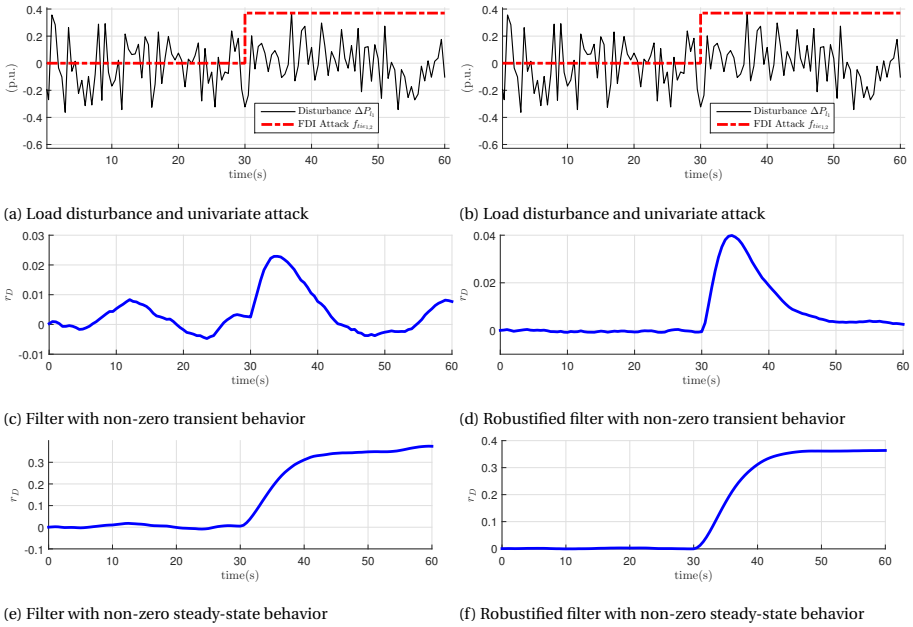
(f) Robustified filter non-zero steady-state behavior

Figure 5.4: Diagnosis filters with or without robustification to model mismatches under univariate attacks and a step-load event. The sampling time $T_s = 0.1$ s, the pole $p = 0.1$, and for $t_s = 10$ s in "training phase", in (5.11), $T = 100$.

disturbance of load deviations (see (4.9) in Section 4.3). Then the mismatch signature $e_x[\cdot]$ can be extracted and further proceed with matrices D_x, G and Q_x. The degree of the residual generator is still set to $d_N = 3$ which is much less than the order of the system dynamics ($n_x = 19$). For a diagnosis filter with stable dynamics, the denominator is set to be in the form $a(q) = (q-p)^{d_N}/(1-p)^{d_N}$ where p is a user-defined variable acting as the pole of the transfer operator $R(q)$, and it is normalized in steady-state value for all feasible poles. This chapter continues to use CPLEX for solving all the corresponding optimization problems.

5.4.2. SIMULATION RESULTS

The first simulation mainly considers the univariate attack scenario that an attacker has manipulated one vulnerable tie-line power flow measurement $\Delta P_{tie_{1,2}}$ at $t = 30$ s in the horizon of 60 s. To challenge the diagnosis filters, a step-load event in Area 1 of the 39-bus system is implemented at $t = 20$ s. In the "training phase" for filter robustification, a simulation time $t_s = 10$ s is chosen and a step-load event occurs at



(a) Load disturbance and univariate attack

(b) Load disturbance and univariate attack

(c) Filter with non-zero transient behavior

(d) Robustified filter with non-zero transient behavior

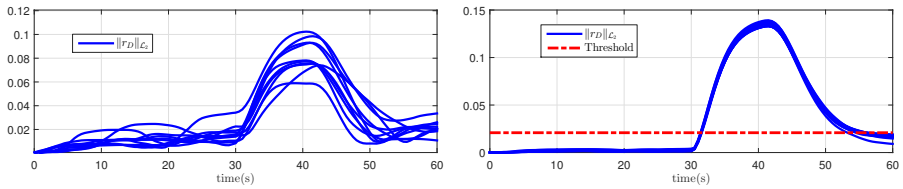
(e) Filter with non-zero steady-state behavior

(f) Robustified filter with non-zero steady-state behavior

Figure 5.5: Diagnosis filters with or without robustification to model mismatches under univariate attacks and stochastic loads. The sampling time $T_s = 0.5\,\text{s}$, the pole $p = 0.8$, and for $t_s = 10\,\text{s}$ in "training phase", in (5.11), $T = 20$.

$t = 2\,\text{s}$. The design variables $\bar{N}$ of the diagnosis filter with robustification are derived from the program (5.15) and to compare, a diagnosis filter is also computed with the objective function $\max_{\bar{N}\in\mathcal{N}} \|\bar{N}\bar{F}_f\|_\infty$ which can be transformed into finite linear programs. Namely, the latter is not robustified to possible model mismatches. Figure 5.4 shows the results of filters with non-zero transient or steady-state behavior. As show in Figure 5.4c and 5.4d, Figure 5.4e and 5.4f, due to the mismatches between the linearized DAE and the detailed PowerFactory simulation model, the robustified diagnosis filter works effectively while the filter without robustification may fail with possible "false alarms". Indeed, the effects of model mismatches starts from the time of step-load event. Besides, from Figure 5.4f we can also see that the robustified diagnosis filter with designed non-zero steady-state behavior from Lemma 4.2.3 can track the value of a univariate attack.

To further challenge the diagnosis filters, this chapter work also considers stochastic load patterns to capture its uncertainty. A number of 100 load disturbance instances are generated for the "training phase" where for each load disturbance without attacks the simulations ($t_s = 10\,\text{s}$) of DAE and PowerFactory are conducted indi-



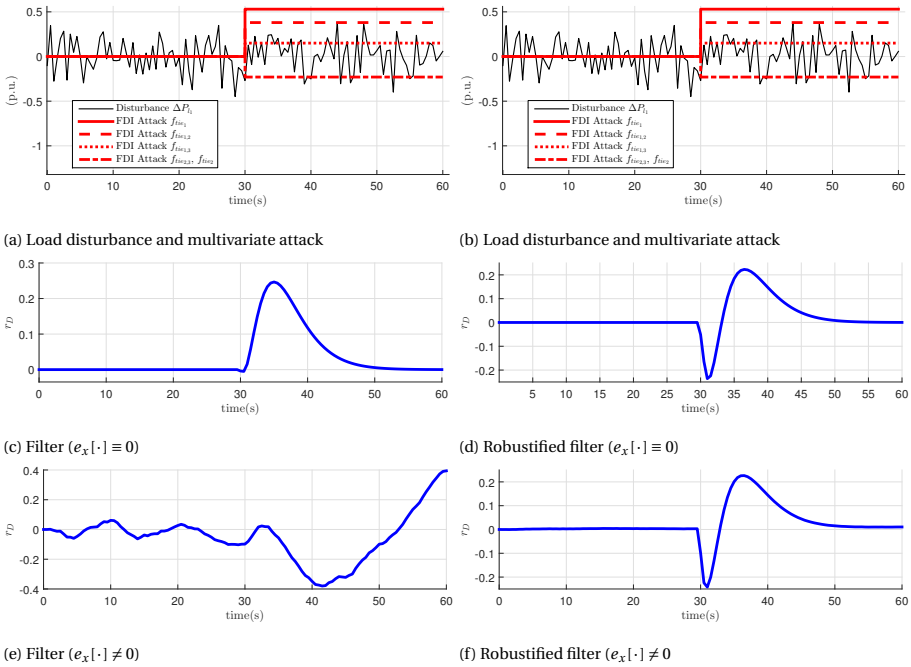
(a) Filter with non-zero transient behavior

(b) Robustified filter with non-zero transient behavior

Figure 5.6: Diagnosis filters with or without robustification to model mismatches under univariate attacks and stochastic loads. The sampling time $T_s = 0.5\,\text{s}$, the pole $p = 0.8$, and for $t_s = 10\,\text{s}$ in "training phase", in (5.11), $T = 20$.

vidually to calculate the model mismatch signatures. In the "test phase", PowerFactory simulations are conducted and the load disturbance has the same stochastic pattern with the ones in the "training phase" and a univariate attack on $\Delta P_{tie_{1,2}}$ at $t = 30\,\text{s}$ is also implemented. Figure 5.5 demonstrates the simulation results. We can see that the robustified diagnosis filter has significant improvements in the regards of mitigating the effects from the model mismatches. Besides, it can approximate the attack value in the non-zero steady-state behavior. For a more clear illustration, Figure 5.6 provides the results of 10 different realizations of load disturbance as well as a univariate attack on $\Delta P_{tie_{1,2}}$. Figure 5.6 depicts the "energy" of the residual signal for the last $t_s = 10\,\text{s}$ under these 10 different instances of load disturbance, namely $\|r_D\|_{\mathcal{L}_2}[\cdot]$. Note that in Figure 5.6b the threshold is set to $\tau^\star + 0.025$, where the square of $\tau^\star$ is the value of $\max_i(\bar{N}Q_{x_i}\bar{N})$ in the 100 training instances ($i \in \{1, \cdots, 100\}$; see the equation (5.14) in Definition 5.3.1), and the added value is computed to avoid possible false alarms, according to [128].

In the second simulation this study moves to the multivariate attack scenario which has been investigated in the previous case study of Chapter 4. There are 5 vulnerable measurements on the tie-lines between each areas, i.e., $\Delta P_{tie_{1,2}}$, $\Delta P_{tie_{1,3}}$, ΔP_{tie_1}, $\Delta P_{tie_{2,3}}$ and ΔP_{tie_2}, and correspondingly there exist 3 basis vectors in the spanning set: $f_1 = [0.1\ \ 0\ \ 0.1\ \ 0\ \ 0]^T$, $f_2 = [0.1\ \ 0.15\ \ 0.25\ \ 0\ \ 0]^T$, $f_3 = [0\ \ 0\ \ 0\ \ 0.1\ \ 0.1]^T$ (all in p.u.). Recall Definition 4.2.5 for characterizing the set of disruptive multivariate attacks. Besides, without loss of generality, the parameters are set to $A = \mathbf{1}^\top$ and $b = 1.5$ in $\mathcal{A}$ and $\eta = 10$ in $\mathcal{N}$. Following Algorithm 2, firstly the program (LP_i) in Corollary 4.4.4 is solved. The optimal value achieves maximum for $i = 2$ that $\gamma_2^\star = 300$, which implies that the diagnosis filter can detect all the admissible attacks in the transient behavior. Next, in the "training phase", similar to the first simulation,



(a) Load disturbance and multivariate attack

(b) Load disturbance and multivariate attack

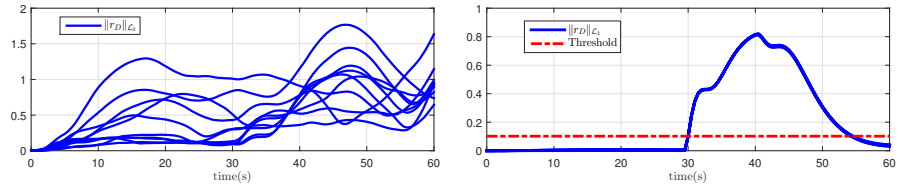
(c) Filter ($e_x[\cdot] \equiv 0$)

(d) Robustified filter ($e_x[\cdot] \equiv 0$)

(e) Filter ($e_x[\cdot] \neq 0$)

(f) Robustified filter ($e_x[\cdot] \neq 0$

Figure 5.7: Diagnosis filters with or without robustification to model mismatches under multivariate attacks and stochastic loads. The sampling time $T_s = 0.5\,\text{s}$, the pole $p = 0.8$, and for $t_s = 10\,\text{s}$ in "training phase", in (5.11), $T = 20$.



(a) Filter with non-zero transient behavior

(b) Robustified filter with non-zero transient behavior

Figure 5.8: Diagnosis filter without or with robustification to model mismatches under univariate attacks and stochastic loads. The sampling time $T_s = 0.5\,\text{s}$ and the pole $p = 0.8$, and for $t_s = 10\,\text{s}$ in "training phase", in (5.11), $T = 20$..

a number of 100 instances of load disturbance are randomly generated, and PowerFactory simulations together with DAE computations in Matlab are performed with the simulation time $t_s = 10\,\text{s}$. After deriving the model mismatch signatures, the program (5.3.7) is solved for the robustified diagnosis filter.

In the "test phase", simulations are conducted that different realizations of load disturbance and multivariate attacks have been implemented. The corruptions are the same with Chapter 4, i.e., $f_{tie_{1,2}} = 0.38\,\text{p.u.}$, $f_{tie_{1,3}} = 0.15\,\text{p.u.}$, $f_{tie_1} = 0.53\,\text{p.u.}$,

$f_{tie_{2,3}} = -0.23\,\text{p.u.}$ and $f_{tie_2} = -0.23\,\text{p.u.}$. The performance of two filters (the robustified filter and the filter without robustification derived from (LP_i) in Corollary 4.4.4 of Chapter 4) is validated with two set of measurements: one measurement set from DAE computations that there is no model mismatch ($e_x[\cdot] \equiv 0$); another set from PowerFactory simulations where the simulation model has mismatches with the DAE model ($e_x[\cdot] \neq 0$). Figure 5.7 demonstrates the simulation results of both diagnosis filters. From Figure 5.7c and 5.7d, we can see that both filter works effectively for the case of $e_x[\cdot] \equiv 0$. However, from Figure 5.7e and 5.7f, when model mismatches exist, the robustified filter still works effectively almost the same to the case of $e_x[\cdot] \equiv 0$, while the filter without robustification totally fails. Note that Figure 5.8 also depicts the "energy" of the residual signal for the last $t_s = 10\,\text{s}$ under 10 different instances of load disturbance. Similarly, the threshold in Figure 5.8b is set to $\tau^\star + 0.1$, where the square of $\tau^\star$ is the value of $\max_i(\bar{N}Q_{x_i}\bar{N})$ in the 100 training instances, and the added value is computed to avoid possible false alarms. Note that for these 5 vulnerable measurements, when looking into the steady-state behavior of the filter with sets $\mathcal{N}$ and $\mathcal{A}$ and solving (4.23) through the programs (4.25a), it turns out that the derived optimal values satisfy the equality $\varphi^\star = \mu^\star = 0$; see Section 4.5.2. This indicates that the optimal multivariate attack with $\alpha^\star$ is a stealthy attack in the long-term horizon, with or without considering the effects from model mismatches. These simulations validate the effectiveness and robustness of the proposed data-assisted model-based diagnosis filter design.

6

Conclusions and Recommendations

This thesis aims for a more cyber-secure intelligent power grid. Motivated by the feasibility of a potent attack that it can be equipped with extensive system knowledge, vast attack resources to manipulate multiple measurements (multivariate attacks) and also strong capability to keep stealthy from possible detectors, the thesis work has built a framework capable of both vulnerability analysis and attack detection. Security index quantifying attack resources was proposed and the attack scenario was extended to subsume the combined attacks. Realistic aspects of limited adversarial knowledge or resources were considered in the overall cyber risk assessment. Co-simulation tool specially for cyber security analysis has been developed, capturing the character of a cyber-physical system of intelligent power grids. A diagnosis filter was designed with a scalable and robust feature to detect all the admissible multivariate attacks by exploiting the attack impact on system dynamics, with non-zero transient or non-zero steady-state residual output. The yielding Nash equilibrium implies that the proposed diagnosis filter is not based on a conservative design in the sense of its long-term behavior. In the end, this thesis also tried to implement the diagnosis filter in a real or simulated power system. A further

robustification method was proposed to mitigate the effects from possible model mismatches on the residual output, assisted by the simulation data to extract the model mismatch signatures.

In the following, a brief summary of this thesis work is presented and some recommendations in the future research are given.

6.1. CONCLUSIONS

This thesis has addressed several research questions, i.e., **Q1** to **Q4** posed in Chapter 1, with an extensive content regarding cyber security of power systems. The main results, including the generality and the limitations of the proposed approaches in each chapter, are concluded as follows.

- **Vulnerability analysis.** In response to **Q1**, Chapter 2 conducted vulnerability assessment of power systems to multivariate data injection attacks and combined attacks. Security index concerning the level of efforts required by the attacker was formulated as a constrained optimization program which was further expressed as a MILP probelm. The combined attacks were compared with multivariate attacks, and for the first time the combined attacks with limited knowledge to the system model were investigated. The results showed that combined attacks can succeed with less attack resources and also expose advantages in keeping stealthy from the typical bad data detection mechanism, bringing more cyber risks to reliable power system operations.

  The approach in Chapter 2 uses the simplified DC power flow model and can be more complex when it comes to the AC model or the system dynamics, but the framework summarized in (1.1) is generic enough to characterize the security index problems in other system models. The resulting MILP formulation is indeed not scalable well and can impose challenges for computation in large-scale power systems. Besides, some aspects of reality are not considered sufficiently in the cyber risk analysis of this thesis. For instance, the attacks may have limited cyber accessibility and can be also "caught" by other repressive measures, while this study assumed that they have been performed successfully when evaluating the "likelihood" of the attacks.

- **Co-simulation for cyber security:** To answer **Q2**, in Chapter 3 this thesis con-

tributes to extend the analytic vulnerability assessment framework to incorporate communication network properties and develop a co-simulation platform to analyze data attacks against the EMS. Methods have been proposed to model measurement routing topologies and the security index formulation was extended to include the communication models. Then the coupling of power system and ICT simulators, including modeling challenges, synchronizations of discrete and continuous simulators and real-time guarantees, were discussed with sufficient details. A co-simulation platform was developed for cyber security experiments. The results show the need of evaluating the vulnerability and attack impact in a comprehensive framework and the possibility to combine system-theoretic and ICT-specific measures to protect power systems from cyber attacks.

The extended vulnerability assessment framework mainly considers the communication topology and data routing schemes. It may face difficulties in further extension to include other properties in the dynamic communication network environment. Besides, for the "simulation" of attacks in OMNeT++, this study only focuses on a direct manipulation of sensors measurements by changing the behavior of the router and the communication link. Not that this has simplified the real intrusion process of a cyber attack.

- **From static to dynamic attack detection:** Responding to **Q3**, the thesis work of Chapter 4 explored the problem of anomaly detection in the power system cyber security with a particular focus on exploiting the dynamics information where tempering multiple measurements data may be possible (multivariate attacks). The study of Chapter 4 showed that a dynamical perspective to the detection task indeed offers powerful diagnosis tools to encounter attack scenarios that may remain stealthy from a static point of view. Two desired features of non-zero transient and non-zero steady-state behavior of the diagnosis filter in the presence of multivariate attacks were investigated, from which theoretical results were provided for the resulting robust optimization programs. The effectiveness of the developed diagnosis filter was validated by simulations in the three-area IEEE 39-bus system.

  The diagnosis filter design builds up on a very generic perspective. It suits

for power system dynamics models presented in the linear DAE formulations even with high dimensions, and thus is tractable for large-scale power systems. To be noted, this diagnosis filter is designed to handle stationary multivariate attacks and all the theoretical results are obtained based on such an assumption. Besides, this diagnosis tool acts as a repressive measure mainly for attacks on sensors measurements, while for attacks compromising other devices like switches or routers, some advanced intrusion detection tools should be developed with enough robustness and effectiveness.

- **Robust attack detection to model mismatches:** To address **Q4**, a further robustification program for the developed diagnosis filter to possible model mismatches has been proposed in Chapter 5. In the end, this thesis aims to implement the developed diagnosis filter in a real or simulated power system where model mismatches could effect the residual output significantly. For this purpose, assisted by the simulation data, the model mismatch signatures could be extracted. Unlike the existing work using pure model-based or data-driven approaches, a novel data-assisted model-based diagnosis filter was proposed and further characterized in an optimization framework to detect univariate attacks and multivariate attacks with non-zero transient or non-zero steady-state residual, while keeping the effects from model mismatches minimized. Numerical results illustrated the effectiveness of the robustified diagnosis filter implemented in the simulator DIgSILENT PowerFactory.

  The proposed robustification scheme provides a generic solution when an operator may have part of the knowledge of the system model to design an anomaly detector but can access the data from the real plant or simulations. To conclude, the diagnosis filter developed in Chapter 4 and Chapter 5 were designed to detect plausible disruptive attacks in an admissible set. The effectiveness of this filter in detecting "basic attacks" which can be also due to sensor or network errors is shown in Figure 4.3 of Chapter 4. The diagnosis filter is also improved to isolate each intrusion and even recover the attack value in Section Chapter 5. However, how to differentiate between attacks, sensors errors and devices failures after a successful detection is not considered in the current diagnosis filter, while these anomalies do have different patterns.

6.2. RECOMMENDATIONS FOR FUTURE WORK

There are several research directions that can be extended from this thesis work. This section introduces some of them.

- **Vulnerability analysis for other system models:** As discussed above, the security index formulation in Chapter 2 is based on linear models in the context of DC state estimation. It would naturally be more complex to compute under the non-linear AC models. With nonlinear (mainly quadratic) constraints and the associated ℓ_0-norm for the objective, the security index problem would be non-convex and difficult to compute. Thus developing convexification or relaxation methods to compute or approximate the security index in nonlinear settings is one of the directions. Besides, considering large-scale power systems, the algorithms should be both effective and computational efficient. For linear dynamic systems, the work in [153] have proposed dynamical security index. Extensions are still needed to include model uncertainties and nonlinearities in the framework. Besides, in the future vulnerability analysis, the "interactions" between the attacks and the preventive or repressive measures should be treated as a dynamic process. Cyber accessibility and the possibility to be "caught" need to be considered from the perspective of the attacker.

- **Cyber attack impact evaluation:** Chapter 2 and Chapter 3 mainly considered the impact of combined attacks and multivariate attacks on the power system in steady-state, and the attack impact on load estimates was formulated for the cyber risk analysis. In fact, attack impact evaluation is complex in general, especially taking into account all the possible cascading events caused by attacks. Besides, some attack impact on economic or physic operations can be difficult to quantify. In the security index formulation of this thesis, the attack impact has not been modeled in the framework while an adversary may also aim to achieve maximum damages in addition to remain stealthy. This is a relevant research direction. Attack impact on system dynamics, such as the consequences on the safety considerations of each state, should also be assessed in the integrated framework of vulnerability and impact analysis.

- **Co-simulation of intelligent power systems:** The co-simulation tool in Chapter 3 is based on the integration of power system and ICT infrastructure sim-

ulators. Although extensive efforts in the literature have been made on co-simulation of hybrid systems, there are still some topics that require further research [68]. How to combine all the heterogeneous models (e.g., statistical models, physical models, or others that provide valuable information) in a scalable way is among one of them. Besides, to support cyber security analysis, the network simulators should have the capability of modeling ICT-specific measures (e.g., intrusion detection system, authentication, etc.) and also diverse attack scenarios. To the best knowledge of this thesis, there is still no user-friendly "attack library" in the simulator level for that purpose. For a better study of the attack behavior, the future research can focus on introducing malicious traffic flows inside the communication environment of a simulator, or compromising a particular communication session to manipulate the specific data. This would also facilitate the study on the the hidden patterns of the ICT-specific intrusion detection/prevention measures, which will be detailed in the next recommendation.

- **Combining system-theoretic and ICT-specific measures:** Throughout the thesis, the detection schemes of bad data or missing data detection mechanism in Chapter 2 and the developed diagnosis filter with a scalable and robust design in Chapter 4 and 5 are all system-theoretic measures for attacks on sensors measurements. This study mainly focuses on the mathematical descriptions of the physical system, while indeed, the "cyber" part of the intelligent power grids, is not discussed sufficiently. From a viewpoint of cyber-physical system, in the future research, the adversarial scenarios should also include cyber-physical attacks on other devices (e.g., switches and routers), and particularly the defense actions should combine system-theoretic and ICT-specific measures [72]. For instance, availability attacks like DoS attacks could also trigger alerts on ICT-specific measures (e.g., intrusion detection system) in addition to the system-theoretic detectors. These features give the opportunities to develop better cross-domain detection mechanisms, while the co-simulation tool from above can support such an analysis, development and validation.

- **Detection of time-varying FDI attacks:** The multivariate false data injection attacks considered in this thesis are stationary according to Assumption 4.2.4,

and as illustrated in Remark 4.4.5 of Chapter 4, the time-varying attacks in the conservative setting that the attacker knows the system model and also the diagnosis filter parameters do impose challenges in the detection tasks. This data attack may bypass any linear residual generator if it is able to dynamically adapt the attack values with full system knowledge. Diagnosis tools are required to address the detection of such type of attacks while ensuring a certain level of robustness, computational efficiency and also practicality.

- **Residual evaluation and anomaly identification:** A residual generator approach for the diagnosis filter is proposed in Chapter 4 and 5. In fact, this thesis put a particular focus on the residual generation in order to achieve the robust detection of all admissible multivariate attacks even there exist model mismatches. The next step is to evaluate the residual and propose a methodology for the threshold computation. In fact, as pointed out by [116], there are few studies on the residual evaluation and threshold computation in the literature work, while such step is key for a decision maker. Some properties such as the statistical behavior of the residual generator when there is no attack but with different disturbances or measurement/process errors could be leveraged to determine the threshold probably. Besides, as pointed out in the previous section, to differentiate between attacks, sensors errors and devices failures, a further anomaly identification scheme needs to be proposed, which can be based on the results of anomaly detection, isolation and recovery of this thesis. In general, cyber attacks may behave in a coordinated manner with specific targets in its kill chain, while errors or failures are more random in nature, and these different patterns can help in anomaly identification.

- **Artificial Intelligence for cyber security:** As discussed, this thesis has tried every effort to implement the developed diagnosis filter in a real or simulated power system, in Chapter 5. A robustification scheme was presented for mitigating the effects of model mismatches on the filter residual with the assistance of simulation data, which would bridge the gaps between model-based and data-driven methods. Other effective approaches may come from combined data-driven and model-based ones. Recent developments in Artificial Intelligence and sensing technology have also provided tools to deal with at-

tacks in high-dimensional, nonlinear and complex power systems. Notably, the reinforcement learning algorithms show some specific advantages. From the perspective of vulnerability analysis, the worst-case scenario may become that an attacker uses reinforcement learning to achieve his targets without a prior knowledge of the system. From the viewpoint of attack detection, a diagnosis tool based on reinforcement learning may also work effectively without a knowledge of the explicit attack models.

BIBLIOGRAPHY

REFERENCES

[1] S. Gorman, "Electricity grid in US penetrated by spies," *The Wall Street Journal*, vol. 8, 2009.

[2] A. Giani, S. Sastry, K. H. Johansson, and H. Sandberg, "The viking project:an initiative on resilient control of power networks," in *2nd International Symposium on Resilient Control Systems*, pp. 31–35, 2009.

[3] INL, "Vulnerability analysis of energy delivery control systems," Tech. Rep. INL/EXT-10-18381, Idaho National Laboratory, Idaho Falls, Idaho, 2011. Available at https://www.energy.gov/oe/downloads/vulnerability-analysis-energy-delivery-control-systems-2011.

[4] J. Hong, Y. Chen, C.-C. Liu, and M. Govindarasu, "Cyber-physical security testbed for substations in a power grid," in *Cyber Physical Systems Approach to Smart Electric Power Grid*, pp. 261–301, Springer, 2015.

[5] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca, R. Schulz, A. Stankovic, C. Taylor, and V. Vittal, "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance," *IEEE Transactions on Power Systems*, vol. 20, pp. 1922–1928, Nov. 2005.

[6] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[7] U.S. GAO, "Critical infrastructure protection: Challenges and efforts to secure control systems," Tech. Rep. GAO-04-628T, United States General Accounting Office, Mar. 2004. Available at https://www.gao.gov/products/GAO-04-628T.

[8] T. M. Chen and S. Abu-Nimeh, "Lessons from stuxnet," *Computer*, vol. 44, no. 4, pp. 91–93, 2011.

[9] N. Falliere, L. O. Murchu, and E. Chien, *W32.Stuxnet Dossier*. Symantec, Feb. 2011. Available at https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf.

[10] B. Kesler, "The vulnerability of nuclear facilities to cyber attack," *Strategic Insights*, vol. 10, no. 1, pp. 15–25, 2011.

[11] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, pp. 3317–3318, July 2017.

[12] D. Case, *Analysis of the cyber attack on the Ukrainian power grid*, Mar. 2016. Available at https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf.

[13] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *16th ACM Conference on Computer and Communications Security*, (New York), pp. 21–32, 2009.

[14] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1362–1370, Sept. 2012.

[15] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems*, (Stockholm), 2010.

[16] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems*, vol. 35, no. 1, pp. 24–45, 2015.

[17] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *49th IEEE Conference on Decision and Control*, pp. 5991–5998, 2010.

[18] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Computer Networks*, vol. 57, no. 5, pp. 1344–1371, 2013.

[19] D. Deka, R. Baldick, and S. Vishwanath, "Optimal data attacks on power grids: Leveraging detection measurement jamming," in *IEEE International Conference on Smart Grid Communications*, (Miami Florida), pp. 392–397, Nov. 2015.

[20] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *IEEE Global Communications Conference*, pp. 3153–3158, 2012.

[21] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, "Data attacks on power system state estimation: Limited adversarial knowledge vs. limited attack resources," in *43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 4313–4318, Oct. 2017.

[22] M. Stifter, J. H. Kazmi, F. Andrén, and T. Strasser, "Co-simulation of power systems, communication and controls," in *IEEE Workshop on Modeling and Simulation of Cyber-Physical Energy Systems*, pp. 1–6, 2014.

[23] M. Wei and W. Wang, "Greenbench: A benchmark for observing power grid vulnerability under data-centric threats," in *The 33rd Annual IEEE International Conference on Computer Communications*, (Toronto), pp. 2625–2633, Apr. 2014.

[24] M. Mallouhi, Y. Al-Nashif, D. Cox, T. Chadaga, and S. Hariri, "A testbed for analyzing security of SCADA control systems (TASSCS)," in *IEEE PES Innovative Smart Grid Technologies Conference*, (Anaheim), Jan. 2011.

[25] M. Findrik, P. Smith, J. H. Kazmi, M. Faschang, and F. Kupzog, "Towards secure and resilient networked power distribution grids: Process and tool adoption," in *IEEE International Conference on Smart Grid Communications*, (Sydney), pp. 435–440, Nov. 2016.

[26] M. P. Barrett, "Framework for improving critical infrastructure cybersecurity," tech. rep., National Institute of Standards and Technology, Apr. 2018. Available at https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf.

[27] J. van den Berg, "Grasping cybersecurity: A set of essential mental models," in *European Conference on Cyber Warfare and Security*, pp. 534–XX, Academic Conferences International Limited, 2019.

[28] E. E. Tiniou, P. Mohajerin Esfahani, and J. Lygeros, "Fault detection with discrete-time measurements: An application for the cyber security of power networks," in *52nd IEEE Conference on Decision and Control*, pp. 194–199, Dec. 2013.

[29] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[30] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, 2015.

[31] J. Zhao, L. Mili, and M. Wang, "A generalized false data injection attacks against power system nonlinear state estimator and countermeasures," *IEEE Transactions on Power Systems*, vol. 33, pp. 4868–4877, Sept. 2018.

[32] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Transactions on Smart Grid*, vol. 9, pp. 1636–1646, May 2018.

[33] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, pp. 1773–1786, Aug. 2016.

[34] T. Wei, X. Chen, X. Li, and Q. Zhu, "Model-based and data-driven approaches for building automation and control," in *IEEE/ACM International Conference on Computer-Aided Design*, pp. 1–8, Nov. 2018.

[35] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, "Cyber risk analysis of combined data attacks against power system state estimation," *IEEE Transactions on Smart Grid*, vol. 10, pp. 3044–3056, May 2019.

[36] K. Pan, A. M. H. Teixeira, M. Cvetkovic, and P. Palensky, "Combined data integrity and availability attacks on state estimation in cyber-physical power grids," in *IEEE International Conference on Smart Grid Communications*, (Sydney), pp. 271–277, Nov. 2016.

[37] NIST, "Guide for conducting risk assessments," Tech. Rep. SP 800-30 R1, National Institute of Standards and Technology, 2012. Available at https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf.

[38] C. Chu and H. H. Iu, "Complex networks theory for modern smart grid applications: A survey," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, pp. 177–191, June 2017.

[39] A. Dwivedi and X. Yu, "A maximum-flow-based complex network approach for power system vulnerability analysis," *IEEE Transactions on Industrial Informatics*, vol. 9, pp. 81–88, Feb. 2013.

[40] G. Chen, Z. Y. Dong, D. J. Hill, G. H. Zhang, and K. Q. Hua, "Attack structural vulnerability of power grids: A hybrid approach based on complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, pp. 595–603, feb 2010.

[41] X. Wei, J. Zhao, T. Huang, and E. Bompard, "A novel cascading faults graph based transmission network vulnerability assessment method," *IEEE transactions on power systems*, vol. 33, no. 3, pp. 2995–3000, 2017.

[42] A. Teixeira, G. Dán, H. Sandberg, and K. H. Johansson, "A cyber security study of a SCADA energy management system: Stealthy deception attacks on the state estimator," *IFAC Proceedings Volumes*, vol. 44, pp. 11271–11277, jan 2011.

[43] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, 2011.

[44] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Transactions on Power Systems*, vol. 29, pp. 627–636, Mar. 2014.

[45] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, pp. 3864–3872, Sept. 2016.

[46] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Transactions on Smart Grid*, vol. 9, pp. 1636–1646, May 2018.

[47] O. Vukovic, K. C. Sou, G. Dan, and H. Sandberg, "Network-aware mitigation of data integrity attacks on power system state estimation," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 6, pp. 1108–1118, 2012.

[48] X. Liu and Z. Li, "Local load redistribution attacks in power systems with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 5, pp. 1665–1676, July 2014.

[49] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1686–1696, 2015.

[50] X. Liu and Z. Li, "False data attacks against AC state estimation with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 8, pp. 2239–2248, Sept. 2017.

[51] M. Esmalifalak, H. Nguyen, R. Zheng, and Z. Han, "Stealth false data injection using independent component analysis in smart grid," in *IEEE International Conference on Smart Grid Communications*, pp. 244–248, Oct. 2011.

[52] J. Kim, L. Tong, and R. J. Thomas, "Data framing attack on state estimation," *IEEE Journal on Selected Areas in Communications*, vol. 32, July 2014.

[53] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber –physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, Jan. 2012.

[54] "Final report on the august 14 2003 blackout in the united states and canada: Causes and recommendations," tech. rep., U.S.-Canada Power System Outage Task Force, Apr. 2004.

[55] A. Teixeira, *Toward cyber-secure and resilient networked control systems*. PhD thesis, KTH Royal Institute of Technology, 2014.

[56] D. Jones, "Statistical analysis of empirical models fitted by optimization," *Biometrika*, pp. 67–88, 1983.

[57] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou, "Efficient computations of a security index for false data attacks in power networks," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3194–3208, 2014.

[58] J. D. Markovic-Petrovic and M. D. Stojanovic, "Analysis of scada system vulnerabilities to DDoS attacks," in *11th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, vol. 02, pp. 591–594, Oct. 2013.

[59] K. C. Sou, H. Sandberg, and K. H. Johansson, "On the exact solution to a smart grid cyber-security analysis problem," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 856–865, 2013.

[60] B. Kang, "Deliverable d4.1-high-level design documentation and deployment architecture for multi-attribute scada intrusion detection system," tech. rep., SPARKS, 2015. project deliverable SPARKS.

[61] M. Hutle, G. Hansch, W. Fitzgerald, T. Hecht, E. Piatkowska, and P. Smith, "Deliverable d2.2-threat and risk assessment methodology," tech. rep., SPARKS, 2015. project deliverable SPARKS.

[62] 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, pp. 1630–1638, July 2017.

[63] G. R. Krumpholz, K. A. Clements, and P. W. Davis, "Power system observability: A practical algorithm using network topology," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, pp. 1534–1542, July 1980.

[64] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid," *Proceedings of the IEEE*, vol. 105, pp. 1389–1407, July 2017.

[65] A. Teixeira, H. Sandberg, G. Dan, and K. H. Johansson, "Optimal power flow: Closing the loop over corrupted data," in *American Control Conference*, pp. 3534–3540, June 2012.

[66] K. Pan, A. Teixeira, C. D. López, and P. Palensky, "Co-simulation for cyber security analysis: Data attacks against energy management system," in *IEEE International Conference on Smart Grid Communications*, pp. 253–258, 2017.

[67] P. Palensky, A. A. Van Der Meer, C. D. Lopez, A. Joseph, and K. Pan, "Cosimulation of intelligent power systems: Fundamentals, software architecture, numerics, and coupling," *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 34–50, 2017.

[68] P. Palensky, A. van der Meer, C. Lopez, A. Joseph, and K. Pan, "Applied cosimulation of intelligent power systems: Implementing hybrid simulators for complex power systems," *IEEE Industrial Electronics Magazine*, vol. 11, pp. 6–21, June 2017.

[69] H. Georg, S. C. Müller, C. Rehtanz, and C. Wietfeld, "Analyzing cyber-physical energy systems:the inspire cosimulation of power and ICT systems using hla," *IEEE Transactions on Industrial Informatics*, vol. 10, pp. 2364–2373, Nov. 2014.

[70] C. B. Vellaithurai, S. S. Biswas, R. Liu, and A. Srivastava, "Real time modeling and simulation of cyber-power system," in *Cyber Physical Systems Approach to Smart Electric Power Grid*, pp. 43–74, Springer-Verlag Berlin Heidelberg, 2015.

[71] K. R. Davis, C. M. Davis, S. A. Zonouz, R. B. Bobba, R. Berthier, L. Garcia, and P. W. Sauer, "A cyber-physical modeling and assessment framework for power grid infrastructures," *IEEE Transactions on Smart Grid*, vol. 6, pp. 2464–2475, Sep 2015.

[72] IEC, "Iec tr 62351-12:2016 power systems management and associated information exchange - data and communications security - part 12: Resilience and security recommendations for power systems with distributed energy resources (der) cyber-physical systems," tech. rep., International Electrotechnical Commission (IEC), 2016. accessed August 2019.

[73] K. Hopkinson, X. Wang, R. Giovanini, J. Thorp, K. Birman, and D. Coury, "EPOCHS: A platform for agent-based electric power and communication simulation built from commercial off-the-shelf components," *IEEE Transactions on Power Systems*, vol. 21, pp. 548–558, May 2006.

[74] D. Bhor, K. Angappan, and K. M. Sivalingam, "A co-simulation framework for smart grid wide-area monitoring networks," in *Sixth International Conference on Communication Systems and Networks*, pp. 1–8, Jan. 2014.

[75] J. Nutaro, P. T. Kuruganti, L. Miller, S. Mullen, and M. Shankar, "Integrated hybrid-simulation of electric power and communications systems," in *IEEE Power Engineering Society General Meeting*, pp. 1–8, June 2007.

[76] H. Lin, S. Veda, S. Shukla, L. Mili, and J. Thorp, "GECO: Global event-driven co-simulation framework for interconnected power system and communication network," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1444–1456, May 2012.

[77] V. Liberatore and A. Al-Hammouri, "Smart grid communication and co-simulation," in *IEEE Energytech*, (Cleveland), pp. 1–5, May 2011.

[78] W. Li, A. Monti, M. Luo, and R. A. Dougal, "VPNET: A co-simulation framework for analyzing communication channel effects on power systems," in *IEEE Electric Ship Technologies Symposium*, (Alexandria), pp. 143–149, 2011.

[79] T. Godfrey, S. Mullen, R. C. Dugan, C. Rodine, D. W. Griffith, and N. Golmie, "Modeling smart grid applications with co-simulation," in *IEEE International Conference on Smart Grid Communications*, pp. 291–296, 2010.

[80] H. Lin, S. Sambamoorthy, S. Shukla, J. Thorp, and L. Mili, "A study of communication and power system infrastructure interdependence on PMU-based wide area monitoring and protection," in *IEEE Power & Energy Society General Meeting*, (San Diego), pp. 1–7, July 2012.

[81] R. Bottura, A. Borghetti, F. Napolitano, and C. A. Nucci, "ICT-power co-simulation platform for the analysis of communication-based volt/var optimization in distribution feeders," in *IEEE PES Innovative Smart Grid Technologies Conference*, (Washington DC), pp. 1–5, Feb. 2014.

[82] M. Armendariz, M. Chenine, L. Nordstrom, and A. Al-Hammouri, "A co-simulation platform for medium/low voltage monitoring and control applications," in *IEEE PES Innovative Smart Grid Technologies Conference*, (Washington DC), pp. 1–5, Feb. 2014.

[83] C. Dufour and J. Belanger, "On the use of real-time simulation technology in smart grid research and development," *IEEE Transactions on Industrial Applications*, vol. 50, pp. 3963–3970, Apr. 2014.

[84] D. Burnier de Castro, S. Ubermasser, S. Henein, M. Stifter, J. Stockl, and S. Hoglinger, "Dynamic co-simulation of agent-based controlled charging electric vehicles and their impacts on low-voltage networks," in *IEEE International Workshop on Intelligent Energy Systems*, (Vienna), pp. 82–88, Nov. 2013.

[85] M. Cvetković, K. Pan, C. David López, R. Bhandia, and P. Palensky, "Co-simulation aspects for energy systems with high penetration of distributed energy resources," in *2017 AEIT International Annual Conference*, pp. 1–6, Sep. 2017.

[86] J. H. Kazmi, A. Latif, I. Ahmad, P. Palensky, and W. Gawlik, "A flexible smart grid co-simulation environment for cyber-physical interdependence analysis," in *Workshop on Modeling and Simulation of Cyber-Physical Energy Systems*, (Vienna), pp. 1–6, Apr. 2016.

[87] I. Ahmad, J. H. Kazmi, M. Shahzad, P. Palensky, and W. Gawlik, "Co-simulation framework based on power system, AI and communication tools for evaluating smart grid applications," in *IEEE PES Innovative Smart Grid Technologies Conference - Asia*, (Bangkok), pp. 1–6, Nov 2015.

[88] C. Yang, G. Zhabelova, C. Yang, and V. Vyatkin, "Cosimulation environment for event-driven distributed controls of smart grid," *IEEE Transactions on Industrial Informatics*, vol. 9, pp. 1423–1435, Aug. 2013.

[89] K. Barnes and B. Johnson, "National scada test bed substation automation evaluation report," tech. rep., Idaho National Laboratory (INL), 2009.

[90] M. Hossain and D. Semere, "Virtual control system development platform with the application of plc device," in *International MultiConference of Engineers and Computer Scientists*, vol. 2, pp. 13–15, 2013.

[91] D. C. Bergman, D. Jin, D. M. Nicol, and T. Yardley, "The virtual power system testbed and inter-testbed integration," in *2Nd Conference on Cyber Security Experimentation and Test*, CSET'09, pp. 5–5, 2009.

[92] R. Liu, C. Vellaithurai, S. S. Biswas, T. T. Gamage, and A. K. Srivastava, "Analyzing the cyber-physical impact of cyber events on the power grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2444–2453, 2015.

[93] C. B. Vellaithurai, S. S. Biswas, R. Liu, and A. Srivastava, "Real time modeling and simulation of cyber-power system," in *Cyber Physical Systems Approach to Smart Electric Power Grid*, pp. 43–74, Springer, 2015.

[94] H. L. Vangheluwe, "DEVS as a common denominator for multi-formalism hybrid systems modelling," in *IEEE International Symposium on Computer-Aided Control System Design*, (Anchorage), pp. 129–134, Sept. 2000.

[95] M. S. Branicky, V. Liberatore, and S. M. Phillips, "Networked control system co-simulation for co-design," in *American Control Conference*, vol. 4, (Denver), pp. 3341–3346, June 2003.

[96] D. Henriksson, A. Cervin, and K.-E. Årzén, "Truetime: Simulation of control loops under shared computer resources," in *15th IFAC World Congress on Automatic Control*, (Barcelona), July 2002.

[97] K. Pan, D. Gusain, and P. Palensky, "Modelica-supported attack impact evaluation in cyber physical energy system," in *IEEE International Symposim on High Assurance Systems Engineering*, pp. 228–233, Jan. 2019.

[98] H. Georg, S. C. Muller, N. Dorsch, C. Rehtanz, and C. Wietfeld, "INSPIRE: Integrated co-simulation of power and ICT systems for real-time evaluation," in *IEEE International Conference on Smart Grid Communications*, (Vancouver), pp. 576–581, Oct. 2013.

[99] W. Li and X. Zhang, "Simulation of the smart grid communications: Challenges, techniques, and future trends," *Computers & Electrical Engineering*, vol. 40, pp. 270–288, Jan. 2014.

[100] R. Bottura, D. Babazadeh, K. Zhu, A. Borghetti, L. Nordstrom, and C. A. Nucci, "SITL and HLA co-simulation platforms: Tools for analysis of the integrated ICT and electric power system," in *IEEE EUROCON*, (Zagreb), pp. 918–925, July 2013.

[101] K. Mets, J. A. Ojea, and C. Develder, "Combining power and communication network simulation for cost-effective smart grid analysis," *IEEE Communications Surveys & Tutorials*, vol. 16, pp. 1771–1796, Mar. 2014.

[102] S. C. Müller, H. Georg, J. J. Nutaro, E. Widl, Y. Deng, P. Palensky, M. U. Awais, M. Chenine, M. Küch, M. Stifter, *et al.*, "Interfacing power system and ict simulators: Challenges, state-of-the-art, and case studies," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 14–24, 2016.

[103] V. Venkataramanan, A. Srivastava, and A. Hahn, "Real-time co-simulation testbed for microgrid cyber-physical analysis," in *Workshop on Modeling and Simulation of Cyber-Physical Energy Systems*, pp. 1–6, Apr. 2016.

[104] D. Babazadeh, M. Chenine, K. Zhu, L. Nordstrom, and A. Al-Hammouri, "A platform for wide area monitoring and control system ICT analysis and development," in *IEEE PowerTech*, (Grenoble), pp. 1–7, June 2013.

[105] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu, "Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid," *IEEE Transactions on Smart Grid*, vol. 4, pp. 847–855, Mar. 2013.

[106] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, pp. 12–19, Feb. 2011.

[107] "Inet framework," 2019. Available at https://inet.omnetpp.org/.

[108] L. Sánchez-Casado, R. A. Rodríguez-Gómez, R. Magán-Carrión, and G. Maciá-Fernández, "Neta: evaluating the effects of network attacks. manets as a case study," in *International Conference on Security of Information and Communication Networks*, pp. 1–10, Springer, 2013.

[109] K. Pan, P. Palensky, and P. M. Esfahani, "From static to dynamic anomaly detection with application to power system cyber security," *IEEE Transactions on Power Systems*, pp. 1–1, 2019.

[110] D. Kirschen and F. Bouffard, "Keeping the lights on and the information flowing," *IEEE Power and Energy Magazine*, vol. 7, pp. 50–60, Jan. 2009.

[111] T. M. Chen and S. Abu-Nimeh, "Lessons from stuxnet," *Computer*, vol. 44, no. 4, pp. 91–93, 2011.

[112] P. Mohajerin Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, "Cyber attack in a two-area power system: Impact identification using reachability," in *American Control Conference*, pp. 962–967, June 2010.

[113] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks," *IEEE Transactions on Smart Grid*, vol. 4, pp. 1244–1253, Sept. 2013.

[114] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, pp. 612–621, Mar. 2014.

[115] M. A. Massoumnia, G. C. Verghese, and A. S. Willsky, "Failure detection and identification," *IEEE Transactions on Automatic Control*, vol. 34, pp. 316–321, Mar. 1989.

[116] S. X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

[117] M. Nyberg and E. Frisk, "Residual generation for fault diagnosis of systems described by linear differential-algebraic equations," *IEEE Transactions on Automatic Control*, vol. 51, pp. 1995–2000, Dec. 2006.

[118] 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, pp. 4760–4774, Sept. 2018.

[119] X. Gao, X. Liu, and J. Han, "Reduced order unknown input observer based distributed fault detection for multi-agent systems," *Journal of the Franklin Institute*, vol. 354, pp. 1464–1483, feb 2017.

[120] P. M. Esfahani and J. Lygeros, "A tractable fault detection and isolation approach for nonlinear systems with probabilistic performance," *IEEE Transactions on Automatic Control*, vol. 61, pp. 633–647, Mar. 2016.

[121] P. Shukla, A. Chakrabortty, and A. Duel-Hallen, "A cyber-security investment game for networked control systems," in *American Control Conference*, pp. 2297–2302, July 2019.

[122] D. Sahabandu, B. Xiao, A. Clark, S. Lee, W. Lee, and R. Poovendran, "Dift games: Dynamic information flow tracking games for advanced persistent threats," in *IEEE Conference on Decision and Control*, pp. 1136–1143, 2018.

[123] D. Sahabandu, S. Moothedath, L. Bushnell, R. Poovendran, J. Aller, W. Lee, and A. Clark, "A game theoretic approach for dynamic information flow tracking with conditional branching," in *American Control Conference*, pp. 2289–2296, July 2019.

[124] R. Deng and H. Liang, "False data injection attacks with limited susceptance information and new countermeasures in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 1619–1628, Mar. 2019.

[125] E. Rakhshani, D. Remon, A. M. Cantarellas, J. M. Garcia, and P. Rodriguez, "Virtual synchronous power strategy for multiple hvdc interconnections of multi-area agc power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1665–1677, 2017.

[126] K. Ogata, *Discrete-time Control Systems (2Nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[127] H. Bevrani, *Robust Power System Frequency Control*. Power Electronics and Power Systems, Springer, 2008.

[128] P. Mohajerin Esfahani, T. Sutter, and J. Lygeros, "Performance bounds for the scenario approach and an extension to a class of non-convex programs," *IEEE Transactions on Automatic Control*, vol. 60, pp. 46–58, Jan. 2015.

[129] 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.

[130] L. Schenato, "To zero or to hold control inputs with lossy links?," *IEEE Transactions on Automatic Control*, vol. 54, pp. 1093–1099, May 2009.

[131] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 911–918, Sep. 2009.

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

[133] W. Ge and C.-Z. FANG, "Detection of faulty components via robust observation," *International Journal of Control*, vol. 47, no. 2, pp. 581–599, 1988.

[134] V. Andrieu and L. Praly, "On the existence of a kazantzis–kravaris/luenberger observer," *SIAM Journal on Control and Optimization*, vol. 45, pp. 432–456, jan 2006.

[135] K. Tidriri, N. Chatti, S. Verron, and T. Tiplica, "Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges," *Annual Reviews in Control*, vol. 42, pp. 63–81, 2016.

[136] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, Nov. 2013.

[137] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, pp. 580–591, Mar. 2014.

[138] X. Luo, Q. Yao, X. Wang, and X. Guan, "Observer-based cyber attack detection and isolation in smart grids," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 127–138, 2018.

[139] H. M. Odendaal and T. Jones, "Actuator fault detection and isolation: An optimised parity space approach," *Control Engineering Practice*, vol. 26, pp. 222–232, 2014.

[140] H. Jiang, M. Xie, and L. Tang, "Markov chain monte carlo methods for parameter estimation of the modified weibull distribution," *Journal of Applied Statistics*, vol. 35, pp. 647–658, jun 2008.

[141] 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, pp. 4985–4995, Sept. 2019.

[142] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online false data injection attack detection with wavelet transform and deep neural networks," *IEEE Transactions on Industrial Informatics*, p. 1, 2018.

[143] A. Ayad, H. E. Z. Farag, A. Youssef, and E. F. El-Saadany, "Detection of false data injection attacks in smart grids using recurrent neural networks," in *IEEE PES Innovative Smart Grid Technologies Conference*, pp. 1–5, Feb. 2018.

[144] Y. Wadhawan, A. AlMajali, and C. Neuman, "A comprehensive analysis of smart grid systems against cyber-physical attacks," *Electronics*, vol. 7, p. 249, oct 2018.

[145] V. B. Krishna, G. A. Weaver, and W. H. Sanders, "PCA-based method for detecting integrity attacks on advanced metering infrastructure," in *Quantitative Evaluation of Systems*, pp. 70–85, Springer International Publishing, 2015.

[146] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Sparse malicious false data injection attacks and defense mechanisms in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 11, pp. 1–12, Oct. 2015.

[147] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Mitigating the impacts of covert cyber attacks in smart grids via reconstruction of measurement data utilizing deep denoising autoencoders," *Energies*, vol. 12, p. 3091, aug 2019.

[148] M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: A reinforcement learning approach," *IEEE Transactions on Smart Grid*, vol. 10, pp. 5174–5185, Sept. 2019.

[149] J. Zhang and A. D. Domínguez-García, "On the impact of communication delays on power system automatic generation control performance," in *North American Power Symposium*, pp. 1–6, Sept. 2014.

[150] R. Doraiswami, "A nonlinear load-frequency control design," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, pp. 1278–1284, July 1978.

[151] R. Ramjug-Ballgobin, S. Z. S. Hassen, and S. Veerapen, "Load frequency control of a nonlinear two-area power system," in *IEEE International Conference on Computing, Communication and Security*, dec 2015.

[152] D. PowerFactory, "39 bus new england system," tech. rep., DIgSILENT GmbH, 2018.

[153] H. Sandberg and A. M. Teixeira, "From control system security indices to attack identifiability," in *Science of Security for Cyber-Physical Systems Workshop*, pp. 1–6, IEEE, 2016.

CURRICULUM VITÆ

Kaikai Pan (潘锴锴) was born on February 20th, 1990 in Pan'an, Zhejiang Province, China. He spent his childhood in a small but beautiful town called Yushan in Pan'an. In 2005, he went to Zhejiang Dongyang High School for his senior high-school study. In 2012, he obtained his B.Sc. degree with honors in measuring and control from Beihang University, Beijing, where he continued his M.Sc. program in instrumentation engineering and got involved in the project of Wind Power Forecasting Using Machine Learning. In 2015, he graduated from Beihang University, being awarded as an Outstanding Graduate of Beijing. During his bachelor and master studies, he has obtained multiple awards and scholarships, including First-class Scholarship for Postgraduate from 2012 to 2015 and National Scholarship in 2011. In 2015, supported by Chinese Scholarship Council, he came to the Intelligent Electrical Power Grids group of Delft University of Technology for his PhD degree. During his four years' PhD study, he was working on cyber security of intelligent power grids, supervised by Prof. dr. Peter Palensky, Dr. Peyman Mohajerin Esfahani and Dr. André Teixeira. He was awarded the IEEE Communications Society Student Travel Grant at the IEEE SmartGridComm conference in Dresden, Germany, in October 2017. He was also one of the prize winners in the PhD poster contest held by Powerweb Institute of TU Delft, in November 2018. His current research interest includes cyber security or resiliency analysis, anomaly detection with model-based or data-driven approaches, cyber-physical systems modeling, Internet-of-Things, machine learning and co-simulation techniques.

List of Publications

Journals:

7. **K. Pan**, A. Teixeira, M. Cvetkovic, & P. Palensky (2018). Cyber Risk Analysis of Combined Data Attacks Against Power System State Estimation. *IEEE Transactions on Smart Grid*, 10(3), 3044–3056. DOI: 10.1109/TSG.2018.2817387

6. **K. Pan**, P. Palensky, & P. Mohajerin Esfahani (2019). From Static to Dynamic Anomaly Detection with Application to Power System Cyber Security. *IEEE Transactions on Power Systems*, PP, 1-1. DOI: 10.1109/TPWRS.2019.2943304

5. P. Palensky, A. van der Meer, C. López, A. Joseph, & **K. Pan** (2017). Applied Cosimulation of Intelligent Power Systems: Implementing Hybrid Simulators for Complex Power Systems. *IEEE Industrial Electronics Magazine*, 11(2), 6–21. DOI: 10.1109/MIE.2017.2671198

4. P. Palensky, A. van der Meer, C. López, A. Joseph, & **K. Pan** (2017). Cosimulation of Intelligent Power Systems: Fundamentals, Software Architecture, Numerics, and Coupling. *IEEE Industrial Electronics Magazine*, 11(1), 34–50. DOI: 10.1109/MIE.2016.2639825

3. **K. Pan**, P. Palensky, & P. Mohajerin Esfahani (2019). Robust Attack Detection in Smart Grid: A Novel Data-assisted Model-based Approach. *To be submitted to IEEE Transactions on Power Systems.*

2. **K. Pan**, E. Rakhshani, & P. Palensky (2019). False Data Injection Attacks on Hybrid AC/HV-DC Interconnected System with Virtual Inertia - Vulnerability, Impact an Detection. *IEEE Transactions on Power Systems, Under Review.*

1. **K. Pan**, Z. Qian, & N. Chen (2015). Probabilistic Short-Term Wind Power Forecasting Using Sparse Bayesian Learning and NWP. *Mathematical Problems in Engineering*, 2015(22), 1–11. DOI: 10.1155/2015/785215

Book Chapter:

1. M. Naglic, A. Joseph, **K. Pan**, M. Popov, M. Meijden, & P. Palensky (2019) Grid Awareness Under Normal Conditions and Cyber-Threats: The PowerWeb Program at TU Delft. In book: Intelligent Integrated Energy Systems, Springer, Cham, 55–75. DOI: 10.1007/978-3-030-00057-8_3

Peer-reviewed Conferences:

6. **K. Pan**, D. Gusain, & P. Palensky (2019) Modelica-Supported Attack Impact Evaluation in Cyber Physical Energy System. In *IEEE 19th International Symposium on High Assurance Systems Engineering (HASE)*, Hangzhou, China, 228-233. DOI: 10.1109/HASE.2019.00042

5. Chenguang Wang, **K. Pan**, Peter Palensky, & Simon Tindemans (2019). Detection of False Data Injection Attacks Using the Autoencoder Approach. *XXI Power Systems Computation Conference (PSCC2020), also as a Special Issue of Electric Power Systems Research, Under Review.*

4. **K. Pan**, A. Teixeira, C. López, & P. Palensky (2017) Co-simulation for Cyber Security Analysis: Data Attacks against Energy Management System. In *8th IEEE International Conference on Smart Grid Communications (SmartGridComm 2017)*, Dresden, Germany, 253-258. DOI: 10.1109/SmartGridComm.2017.8340668

3. **K. Pan**, A. Teixeira, M. Cvetkovic, & P. Palensky (2017) Data Attacks on Power System State Estimation: Limited Adversarial Knowledge vs. Limited Attack Resources. In *43rd Annual Conference of the IEEE Industrial Electronics Society (IECON 2017)*, Beijing, China, 4313-4318. DOI: 10.1109/IECON.2017.8216741

2. M. Cvetkovic, **K. Pan**, C. David López, R. Bhandia, & P. Palensky (2017) Co-simulation Aspects for Energy Systems with High Penetration of Distributed Energy Resources. In *2017 AEIT International Annual Conference*, Cagliari, Italy, 1-6. DOI: 10.23919/AEIT.2017.8240488

1. **K. Pan**, A. Teixeira, M. Cvetkovic, & P. Palensky (2016) Combined Data Integrity and Availability Attacks on State Estimation in Cyber-Physical Power Grids. In *7th IEEE International Conference on Smart Grid Communications (SmartGridComm 2016)*, Sydney, 1-7. DOI: 10.1109/SmartGridComm.2016.7778773

ACKNOWLEDGEMENTS

"At thirty, I had planted my feet firm upon the ground." This quote by Confucius has encouraged me in the past four years. Indeed, doing a PhD is a process of self-improving, and it is this enjoyable experience in TU Delft that makes me feel more ready for my coming "thirty". Now, I would like to express my sincere gratitude to all those who have helped me and been with me during this incredible journey.

First of all, my deepest thankfulness goes to my promotor Prof.dr. Peter Palensky and my copromotor Dr. Peyman Mohajerin Esfahani. Peter gave me the wonderful opportunity to do the PhD research in TU Delft. He supported me to enter the "uncharted terrain" for me - power system cyber security, and I would like to say I enjoyed it so much. His enthusiasm in research, great vision, smart insight, and open-mindedness have not only impressed me but also affected me, and become a continuous source of inspiration for me. Thank you Peter! I also feel fortunate to have Peyman as my copromotor and daily-supervisor. This thesis work can not be carried out smoothly without his supervisions. The mathematical part is written with a huge help from him. We had so many discussions with great contents. I still remember our whole-afternoon talks during weekends in the library while other buildings were closed! He is hard-working, efficient, patient and intelligent that he has set an example for me in my academic career. Thank you Peyman! I hope we can continue our collaborations in the future!

A special thank to Dr. André Teixeira from Uppsala University. He is not only a collaborator or a co-author for me, but also a copromotor and a supervisor in the first two years of my PhD. He helped me to start my PhD research when I was a beginner in the domain of cyber security and provided me with constant supports. The thesis work of vulnerability analysis was supervised by him, which has a big impact on the whole research during my PhD. Thank you André!

Next I would like to thank my PhD committee members, Prof.dr.ir. Jan van den Berg, Prof.dr. Xinghuo Yu, Prof.dr.ir. J.A. la Poutré, Dr. Anurag. K. Srivastava, Dr.ir.

José Rueda Torres and Prof.dr.ir. Miro Zeman for their assessment of this thesis and constructive and insightful comments for me to improve.

I am also thankful to the professionals who have helped me with my research work. Thanks to my colleague and friend Dr. Milos Cvetkovic for his help and recommendations when I was looking for jobs. We had fruitful discussions and good collaborations. Thanks to Prof. Sebastian Lehnhoff for his nice hospitality during my short visit in OFFIS, Oldenburg, and thanks to all the colleagues there for their assistance when I was learning the co-simulation framework Mosaik. A kind thanks to Prof. Ralf Boden, Dr. Daisuke Mashima for the nice talks we had in the SmartGridComm conferences, and to Dr. Jawad Kazmi for his fast responses when I had any questions in using OMNeT++. Thanks to Ibrahim Diab for his help when we were recording the online courses on smart grid basics and modeling.

My gratitude also goes to my master's supervisor Prof. Zheng Qian from Beihang University. Without his supports and help, I would not pursue my PhD degree in TU Delft. We had so many happy moments when I was in his research team!

I feel really happy to work in the IEPG group of the ESE department. A deep thanks to the secretaries Ellen, Sharmila and Carla for their patient and kind help. Thanks to the academic staff Prof.ir. Mart van der Meijden, Dr.ir. Marjan Popov, Dr. Simon Tindemans and Dr. Alex Stefanov. I can always get some suggestions and inspirations from them. Thanks to Elyas, Claudio, Digvijay, Rishabh, Arun and Arjen. They are my nice colleagues, friends and also co-authors. Thanks to Bart Tuinema for checking the Dutch translation of this thesis summary. Thanks to my former and current officemates and colleagues, Mario, Swasti, Romain, Nakisa, Matija, Ilya, Arcadio, Hazem, Vinay, Umer, José, Ebrahim, Zameer. Many thanks to my Chinese friends: Lian Liu, Zhou Liu, Siyuan, Chenguang, Aihui, Haiyan and Da Wang from our group, and Zian Qin, Yang Wu, Yunhe from DCE&S group, and Guangtao, Can Han from PVMD group, and Jingwei from DCSC of 3mE faculty. We had great memories in both academic and social activities.

I want to thank all of my friends during theses few years in Delft. Thanks to Zhang Cao and Xuan for helping me to start my new life in the Netherlands. Thanks to Ding Ding, Shengyue, Hao Yu, Riming, Jie Zhang, Xuerui, Sihao, Yan Liu, Lixue, Shuo Li, Kai Liu, Ziqiao, Shuaiqiang, Jiao Chen, Anqi, Hai Gong, Na Li, etc. We have experienced many things together. Thanks to my friends Zhi Hong, Meng Wang and

the USSR instructors and members. We shared a lot of fun in the badminton courts. Thanks to my roommates Dingsihao Lyu, Maolong, Nianlei, Biling, Aoge, Qingxi and Qingzhu. With them I had the chance to try the local food from different cities!

Finally, I want to express my warmest thanks to my family. Thanks to my parents for their unconditional supports. I promise that, I would have more time with them in the near future. Thanks to my love Chang Chen, who have stood by me and supported me whenever I needed. All I want to say is, I love you! Finally, thanks to my grandfather and grandmother, and wish them good health.

Kaikai Pan
Delft, The Netherlands