KTH Electrical Engineering

Toward Secure and Reliable Networked Control Systems

ANDRÉ TEIXEIRA

Licentiate Thesis
Stockholm, Sweden 2011

KTH School of Electrical Engineering
Automatic Control Lab
SE-100 44 Stockholm
SWEDEN

TRITA-EE 2011:075
ISSN 1653-5146
ISBN 978-91-7501-197-4

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie licentiatexamen i reglerteknik fredagen den 9 december 2011, klockan 09:00 i sal L1, Kungliga Tekniska högskolan, Drottning Kristinas väg 30, Stockholm.

Tryck: Universitetsservice US AB

Abstract

Security and reliability are essential properties in Networked Control Systems (NCS), which are increasingly relevant in several important applications such as the process industry and electric power networks. The trend towards using non-proprietary and pervasive communication and information technology (IT) systems, such as the Internet and wireless communications, may result in NCS being vulnerable to cyber attacks. Traditional IT security does not consider the interdependencies between the physical components and the cyber realm of IT systems. Moreover, the control theoretic approach is not tailored to handle IT threats, focusing instead on nature-driven events. This thesis addresses the security and reliability of NCS, with a particular focus on power system control and supervision, contributing towards establishing a framework capable of analyzing and building NCS security.

In our first contribution, the cyber security of the State Estimator (SE) in power networks is analyzed under malicious sensor data corruption attacks. The set of stealthy attacks bypassing current Bad Data Detector (BDD) schemes is characterized for the nonlinear least squares SE, assuming the attacker has accurate knowledge of a linearized model. This result is then extended to uncertain models using the geometric properties of the SE and BDD. Using the previous results, a security framework based on novel rational attack models is proposed, in which the minimum-effort attack policy is cast as a constrained optimization problem. The optimal attack cost is interpreted as a security metric, which can be used in the design of protective schemes to strengthen security. The features of the proposed framework are illustrated through simulation examples and experiments.

As our second contribution, we analyze the behavior of the Optimal Power Flow (OPF) algorithm in the presence of stealthy sensor data corruption and the resulting consequences to the power network operation. In particular, we characterize the set of attacks that may lead the operator to apply the erroneous OPF recommendation and propose an analytical expression for the optimal solution of a simplified OPF problem with corrupted measurements. A novel impact-aware security metric is proposed based on these results, considering both the impact on the system and the attack cost. A small analytical example and numerical simulations are presented to illustrate and motivate our contributions.

The third contribution considers the design of distributed schemes for fault detection and isolation in large-scale networks of second-order systems. The proposed approach is based on unknown input observers and exploits the networked structure of the system. Conditions are given on what local measurements should be available for the proposed scheme to be feasible. Infeasibility results with respect to available measurements and faults are also provided. In addition, methods to reduce the complexity of the proposed scheme are discussed, thus ensuring the scalability of the solution. Applications to power networks and robotic formations are presented through numerical examples.

Acknowledgements

First of all, I would like to thank Karl H. Johansson and Henrik Sandberg for their guidance and inspiration throughout my journey in academic research, from even before my Master thesis all the way up to this Licentiate thesis. You have shown me the value of good problem formulations and intuition in doing research! Looking forward to the journey yet to come!

I am grateful for the short but fantastic visiting periods spent abroad. Thanks to Saurabh Amin, Annarita Giani, Kameshwar Poolla, and Shankar Sastry for the time I spent at UC Berkeley and the great discussions we had. Thanks to James Anderson and Antonis Papachristodoulou for the wonderful moments at the University of Oxford. It was an intensive, but quite enjoyable research experience!

Special thanks to György Dán and Iman Shames for the helpful discussions and collaborations we had so far. I would also like to thank Sonja Buchegger and Luigi Vanfretti for being in my reference group.

I have had fantastic times at the Automatic Control Lab so far. For that I'm particularly thankful to José Araújo for motivating the beginning of my journey abroad and for his treasured friendship. Many thanks to my current and former colleagues at the lab, in particular Assad Alam, Phoebus Chen, Burak Demirel, Farhad Farokhi, António Gonga Euhanna Ghadimi, Per Hägg, Christian Larsson, Piergiuseppe Di Marco, Pangun Park, Chithrupa Ramesh, Pablo Soldati, Kin Cheong Sou, Christopher Sturk, Ubaldo Tiberi, Jim Weimer, and Zhenhua Zou for all the support and helpful discussions.

I am grateful to the Master's students I supervised, Daniel P. Huertas and Albert F. Navarro, to whom I wish all the best.

This work has been financially supported by the EU FP7 project VIKING, the Swedish Research Council, the Swedish Foundation for Strategic Research and the Swedish Governmental Agency for Innovation Systems. Their support is greatly acknowledged.

None of this would have been possible without the support from my family. My dearest gratitude to my wife Cátia for bravely and successfully facing the challenges of studying and teaching abroad and, most of all, being married to a PhD candidate. Special thanks to my parents Manuel and Maria Eugénia and my brother Daniel for their support and nurturing throughout my life.

André Teixeira

Contents

List of Abbreviations

BDD	Bad Data Detector
CA	Contingency Analysis
EMS	Energy Management System
FDI	Fault Detection and Isolation
IT	Information Technology
NCS	Networked Control Systems
OPF	Optimal Power Flow
RTU	Remote Terminal Unit
SCADA	Supervisory Control and Data Acquisition
SE	State Estimator
UIO	Unknown Input Observer
WACS	Wide-Are Control System
WAMS	Wide-Area Monitoring System
WLS	Weighted Least Squares

Chapter 1

Introduction

Control engineering and automation are essential components in modern societies: from supporting and enhancing industrial processes to ensuring that electricity is continuously provided to every domestic household. The ubiquitous use of automatic control is very much due to the technological developments in computation, actuation, and sensing. Meeting safety, reliability, and performance requirements needs continuous monitoring and control, only feasible through automation. Control theory has contributed to methods for guaranteeing these properties with primary focus on centralized controllers with full access to all the measurements and actuators and reliable sensor-to-controller and controller-to-actuator communications links (Samad and Annaswamy, 2011).

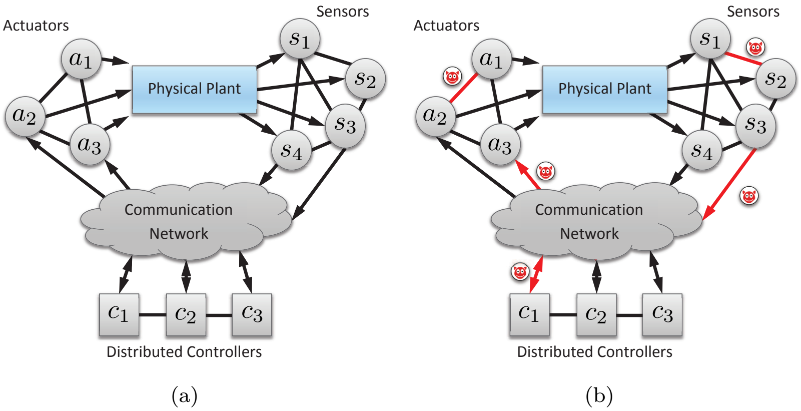


Figure 1.1: A networked control system under (a) nominal behavior and (b) cyber attacks on actuators, sensors, controllers, and communication links.

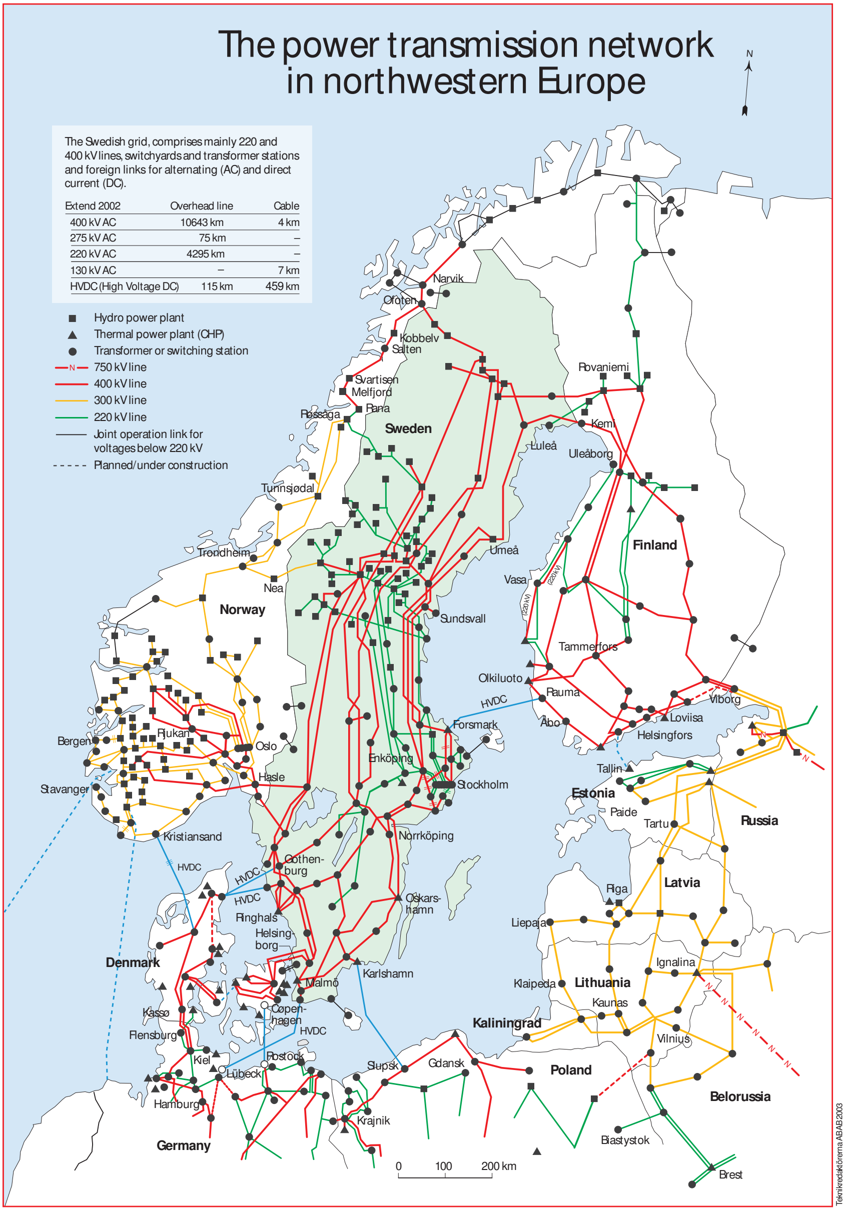
The technological development has led to the increased use of digital controllers and communication networks in many control applications, effectively transforming them into Networked Control Systems (NCS), i.e., systems controlled over communication networks as depicted in Figure 1.1(a) (Samad et al., 2007). Although this paradigm shift creates new opportunities to increase the overall system safety, reliability, and performance, it also leads to new challenges, specially when the effects of the communication network influences the control performance (Baillieul and Antsaklis, 2007; Hespanha et al., 2007). Additional challenges come from the use of open (non-proprietary) and pervasive communication and information technology (IT) systems, such as the Internet and common PC operating systems and wireless communication technologies. As a result, NCS may become vulnerable to cyber threats, as illustrated in Figure 1.1(b) (Bishop, 2002). Traditionally IT security does not consider the interdependencies between the physical components and the IT system. Theory and tools to analyze and build NCS security are therefore lacking and in need to be developed, requiring a systematic handling of the complex coupling between the physical and cyber realms in these systems.

1.1 Motivating Example: Power Networks

To illustrate and motivate the relevance of NCS security, we consider a complex and large-scale application essential to modern society: power transmission networks.

Power transmission networks are complex and spatially distributed systems, as illustrated in Figure 1.2. They are operated through supervisory control and data acquisition (SCADA) systems, which represent the backbone IT and control infrastructure. SCADA systems collect data from remote terminal units (RTUs) installed in substations and relay aggregated measurements to the central master station located at the control center. SCADA systems for power networks are complemented by a set of application specific software, usually called energy management systems (EMS), enabling state and measurement estimation and optimal operation under safety and reliability constraints.

As discussed in (Giani et al., 2009), there are several vulnerabilities in the SCADA system architecture, as illustrated in Figure 1.3. These include the direct tampering of RTUs (A1 and A5), communication links between the RTUs and the control center (A2 and A6), and the IT software and databases in the control center (A4 and A7). There are several reports regarding cyber attacks on SCADA systems for power networks (CBSNews, 2009; Gorman, 2009). In addition, as seen in the U.S.-Canada blackout in August 2003 (U.S.-Canada PSOTF, 2004), the malfunction of SCADA systems together with nature-driven failures may result in major blackouts. Hence one can hypothesize that cyber attacks may have similar impact. On the other hand, in some power networks the supervisory operation is market-driven, meaning that the prices paid to power producers vary according to the current estimated state of the system and the available resources. The California electricity crisis in 2000–2001 (FERC, 2003), a consequence of both a flawed market design and covert market manipulations, shows that there may exist



The Swedish grid, comprises mainly 220 and 400 kV lines, switchyards and transformer stations and foreign links for alternating (AC) and direct current (DC).

Extend 2002	Overhead line	Cable
400 kV AC	10643 km	4 km
275 kV AC	75 km	–
220 kV AC	4295 km	–
130 kV AC	–	7 km
HVDC (High Voltage DC)	115 km	459 km

Figure 1.2: Nordic power network with indications of hydro (■) and thermal (▲) power plants and substations (•) (SvK).

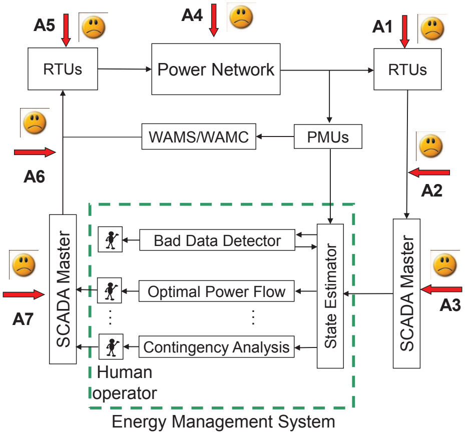


Figure 1.3: A schematic block diagram of a power network, a SCADA system, a control center, and possible IT vulnerabilities.

economic incentive to tamper with the power system operation.

Regarding power generator plants, staged cyber attacks have succeeded in physically damaging generators (Meserve, 2007). More recently, two advanced recent cyber attacks targeting industrial control systems were reported, namely Stuxnet and Duqu (Falliere et al., 2011; Symantec, 2011). Stuxnet was designed to physically damage heavy machinery like steam turbines and gas centrifuges present in process plants by interfering with low-level actuators (Falliere et al., 2011; Rid, 2011). Duqu, the most recent one, seems to be based on Stuxnet's source code and was aimed at espionage attacks on industrial control systems manufacturers in an attempt to obtain sensitive information for facilitating future attacks (Symantec, 2011). Although no hardware damage was confirmed, both these threats highlight the importance of secure IT and NCS systems.

1.2 Problem Formulation

This thesis addresses the problem of NCS security and reliability. Regarding security, this thesis contributes towards a comprehensive framework to analyze, identify, and evaluate the consequences of existing vulnerabilities in NCS, essential for proposing effective protection schemes. As for reliability, monitoring schemes suitable for large-scale systems are proposed.

In particular, the security of supervisory monitoring and control of power transmission networks is analyzed in detail regarding cyber attacks on the measurements. Figure 1.4 illustrates a control center with a SCADA EMS operating a power network. The received measurements are processed by the state estimator (SE) to estimate the state and unmeasured variables of the system. These are used by

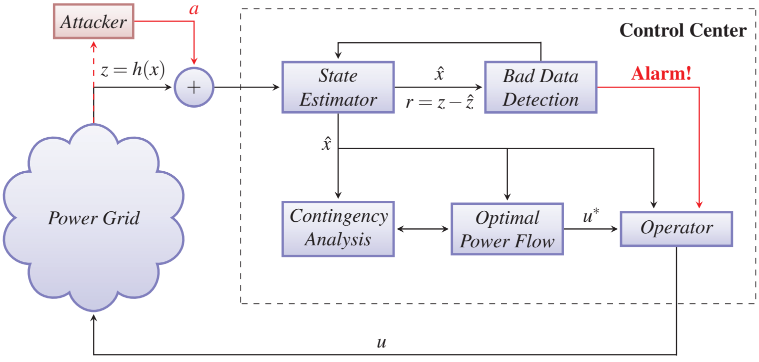


Figure 1.4: The SCADA EMS under a cyber attack on the measurements.

other EMS components to provide recommended control actions to the operator.

In our scenario, the set of measurements z is corrupted by the attacker with a set of additive bias a. Although there exist bad data detectors (BDD), the attack may pass undetected and affect the subsequent EMS components and the operator's decisions. How these attacks affect the system operation depends on the interconnections between the several EMS components and the operator. This thesis analyzes such interdependencies, characterizing the class of attacks that are not detected and studying their effect on the optimal power flow algorithm and power network operation.

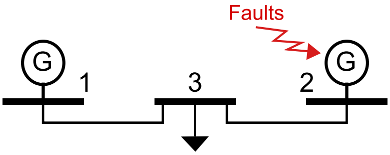


Figure 1.5: A power transmission network with faults in a generator bus.

Reliability of power transmission networks is also addressed. Consider the power network depicted in Figure 1.5 with a faulty generator. Due to the networked structure of the system, faults occurring in a generator can propagate to the entire system, possibly leading to cascading failures of other generators. An approach to meet the reliability constraints is to timely detect, locate, and mitigate possible faults before other components are affected. This thesis focus on the former two, namely detection and localization of faults using model-based approaches. Given the dimension of power transmission networks as illustrated in Figure 1.2, suitable fault monitoring schemes are required to be distributed.

1.3 Outline and Contributions

This thesis is the compilation and edition of results presented or submitted to peer-reviewed scientific venues. In the following we present the outline of the thesis, in addition to the collection of papers the respective chapters are based on.

Chapter 2: Background

The background on NCS and standard IT security is given, followed by a more detailed problem formulation and description of supervisory monitoring and control of power networks.

Chapter 3: Cyber Security of State Estimator in Power Systems

This chapter considerers the current SE and BDD algorithms in SCADA EMS for power networks under malicious measurement data corruption attacks. A novel attacker model is proposed, including reasonable attack goals, constraints, and costs. The set of stealthy attacks bypassing current BDD schemes is characterized for the nonlinear least squares SE, assuming the attacker has accurate knowledge of a linearized model. This result is then extended to uncertain models using the geometric properties of the SE and detection schemes. Using the previous results, a security framework based on the novel attack models is proposed, in which the minimum-effort attack policy is cast as a constrained optimization problem. The optimal attack cost is interpreted as a security metric, which can be used in the design of protective schemes to strengthen security. The features of the proposed framework are illustrated through simulation examples and experiments.

This work is based on the following publications.

- 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, Sweden, 2010.
- A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. Sastry, "Cyber Security Analysis of State Estimators in Electric Power Systems". In Proceedings of the 49th Conference on Decision and Control, Atlanta, GA, USA, 2010.
- A. Teixeira, G. Dán, H. Sandberg, and K. H. Johansson, "Cyber Security Study of a SCADA Energy Management System: Stealthy Deception Attacks on the State Estimator". In 18th IFAC World Congress, Milan, Italy, 2011.

Chapter 4: Cyber-Security of Optimal Power Flow in Power Systems

In this chapter we analyze the behavior of the Optimal Power Flow (OPF) algorithm in the presence of maliciously biased estimates and the resulting increased operation costs to the system operator. In particular, we characterize the set of undetected attacks that may lead the operator to apply the erroneous OPF recommendation. A new impact-aware security metric is introduced using the previous analysis, which is useful for allocation and prioritization of protective measures. Additionally, we propose an analytical expression for the optimal solution of a simplified OPF problem with corrupted measurements. Analytical and numerical examples are discussed to illustrate our contributions.

This work is based on the following paper, which has been recently submitted.

- A. Teixeira, H. Sandberg, G. Dán, and K. H. Johansson. "Optimal Power Flow: Closing the Loop over Corrupted Data". Submitted to the American Control Conference, 2012.

Chapter 5: Distributed Fault Diagnosis in Networked Systems

This chapter considers physical failures and distributed schemes for their detection and isolation. The existence of unknown input observers (UIO) for networks of interconnected second-order linear time invariant systems is studied. Two classes of distributed control systems of large practical relevance are considered. It is proved that for these systems one can construct a bank of UIOs, and use them to detect and isolate faults in the network through a distributed implementation. In particular, by exploiting the system structure, this work provides further insight into the design of UIO for networked systems. Moreover, the importance of certain network measurements is shown. Infeasibility results with respect to available measurements and faults are also provided, as well as methods to remove faulty agents from the network. Applications to power networks and robotic formations are presented. It is shown how the developed methodology apply to a power network described by the swing equation with a faulty bus. For a multi-robot system, it is illustrated how a faulty robot can be detected and removed.

This work is based on the following publications.

- A. Teixeira, H. Sandberg, and K. H. Johansson, "Networked Control Systems under Cyber Attacks with Applications to Power Networks". In American Control Conference, Baltimore, MD, USA, 2010.

- I. Shames, A. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed Fault Detection for Interconnected Second-Order Systems with Applications to Power Networks". In First Workshop on Secure Control Systems, Stockholm, Sweden, 2010.

- I. Shames, A. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed Fault Detection for Interconnected Systems". In Automatica, to appear, 2011

- I. Shames, A. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed Fault Detection and Isolation with Imprecise Network Models". Submitted to the American Control Conference, 2012

Chapter 6: Conclusions and Future Work

A summary of the thesis contributions and future research directions are discussed.

1.4 Other Contributions

The following publications by the author are not covered in this thesis.

- J. Anderson, A. Teixeira, H. Sandberg and A. Papachristodoulou. "Dynamical System Decomposition Using Dissipation Inequalities". To appear in Proceedings of the 50th Conference on Decision and Control and European Control Conference, Orlando, FL, USA, 2011.
- M. Larsson, J. Lindberg, J. Lycke, K. Hansson, A. Khakulov, E. Ringh, F. Svensson, I. Tjernberg, A. Alam, J. Araujo, F. Farokhi, E. Gadhimi, A. Teixeira, D. V. Dimarogonas, K. H. Johansson, "Toward an Indoor Testbed for Mobile Networked Control Systems". Submitted to the First Workshop on Research, Development and Education on Unmanned Aerial Systems (RED-UAS 2011), Seville, Spain 2011.
- I. Shames, A. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed Leader Selection without Direct Inter-Agent Communication". In IFAC Workshop on Distributed Estimation and Control of Networked Systems (NEC-SYS), Annecy, France, 2010.

Chapter 2

Background

In this chapter we give an overview of the traditional frameworks for NCS reliability and safety and IT security relevant to this thesis. The main application considered in this work, power transmission networks, is also described.

2.1 Model-Based Fault Diagnosis

System performance, reliability, and safety concern the ability of maintaining different levels of acceptable behavior in spite of unpredicted events. Performance is mainly addressed by compensating effects of disturbances, while possible approaches to reliability and safety include fault diagnosis and fault tolerant control. This section provides a general overview of model-based fault diagnosis methods (Chen and Patton, 1999; Ding, 2008; Hwang et al., 2010). In particular, we consider fault detection and isolation.

Detection schemes for static models under cyber attacks are studied in Chapter 3, while in Chapter 5 we consider fault detection and isolation methods for large-scale dynamic models.

2.1.1 Model-Based Fault Detection

The objective of fault detection is to assess whether the system is in nominal behavior (no faults), or in an abnormal behavior (with faults). In model-based fault detection, the nominal behavior of the system can be predicted based on known plant models and respective inputs. The basic principle in model-based fault detection is then to compare the predicted and real system trajectories, obtaining the so-called residual as illustrated in Figure 2.1. The system is declared faulty if there is a significant mismatch in the residual signal.

Residual generation

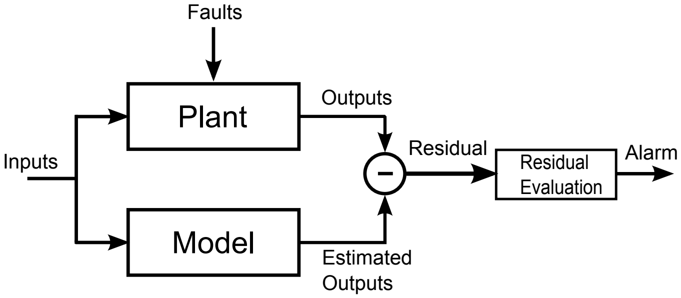


Figure 2.1: Block diagram of a generic model-based fault detection scheme.

Hwang et al. (2010) give an overview of the several approaches to model-based fault detection, isolation, and recovery. Regarding fault detection, one of the main problem is the computation of the residual signal, i.e., a signal quantifying the mismatch between the real and predicted outputs. This is particularly important in the presence of measurement and process noise, unknown disturbances, and model uncertainties. A widely used class of model-based residual generation schemes is the observer-based approach (Patton and Chen, 1997). In this approach an observer is designed to estimate the state and output of the plant, which is then compared to the real plant output to generate the residual.

Example 2.1
Consider the model

$$z = Hx + f = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} x + f,$$

where z is the set of measurements, x is the unknown state, and f is a possible fault. Consider the nominal fault-free case where one has $f = 0$. An observer-based approach to generate a residual is to estimate the state x through linear least-squares, yielding $\hat{x} = (H^\top H)^{-1} H^\top z$. This estimate can then be used to generate the following residual $r = z - H\hat{x} = (I - H(H^\top H)^{-1}H^\top)z$. Note that in the faulty case $f \neq 0$ we have $r = (I - H(H^\top H)^{-1}H^\top)f$. This residual can detect faults f on the measurements only if the faults do not satisfy the model, i.e. $f \notin \operatorname{Im} H$. This will be further discussed in Chapter 3.

The former example illustrated an observer-based method for a linear static system. Similar approaches exist for dynamical systems as well, using for instance full-order observers (Patton and Chen, 1997) or the Kalman filter (Chow and Willsky, 1984). In the presence of additional uncertainties as unknown disturbances, other techniques must be employed. Examples of such techniques include robust observers compensating the disturbance effect (Douglas and Speyer, 1995), optimization-based observer design mitigating the disturbance effects while maximizing the sensitivity to faults (Chung and Speyer, 1998), and unknown input observers (UIO) that are able to completely decouple the disturbance effect from the state estimate (Chen et al., 1996). The UIO approach will be used in Chapter 5.

Example 2.2

Consider now the previous example with a disturbance

$$z = Hx + Bd = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} d,$$

where d is a scalar disturbance. To obtain a residual decoupled from d, one can pre-multiply the measurements by $P = I - B(B^\top B)^{-1}B^\top$, resulting in $w = Pz = PHx + PBd = PHx$. If $\tilde{H} = PH$ is full-column rank, one can compute the disturbance decoupled residual $\tilde{r} = w - \hat{w} = (I - \tilde{H}(\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top)Pz$. The UIO is a dynamic equivalent of the approach taken in this example, suitable for dynamical linear systems.

Residual evaluation

Another important issue in fault detection and isolation is the residual evaluation (Hwang et al., 2010). The objective of this evaluation is to decide whether or not a fault is present for a given residual signal. In deterministic systems, residual evaluation may be performed by comparing the norm of the residual signal against a threshold chosen to ensure robustness to uncertainties (Ding, 2008). In stochastic systems, the statistical model of the residual signal can be used to design optimal evaluation schemes in the form of hypothesis test, for instance the Generalized Likelihood Ratio Test, Sequential Probability Ratio Test, and the CUSUM (Basseville and Nikiforov, 1993; Hwang et al., 2010). The hypotheses test approach is used in Chapter 3.

2.1.2 Model-Based Fault Isolation

In addition to fault detection, it is useful to locate the faulty component in the system, which is called fault isolation (Ding, 2008; Hwang et al., 2010). Fault isolation is usually a harder problem than fault detection and may require additional model knowledge. Since for fault isolation one needs to distinguish between different possible faults, the model of the system in different faulty conditions is required, see Figure 2.2.

A common approach is to constrain the design of the residual generator so that the residuals have a certain structure facilitating isolation. Possible methods include the Beard-Jones filter, designed so that each fault excites the residual in

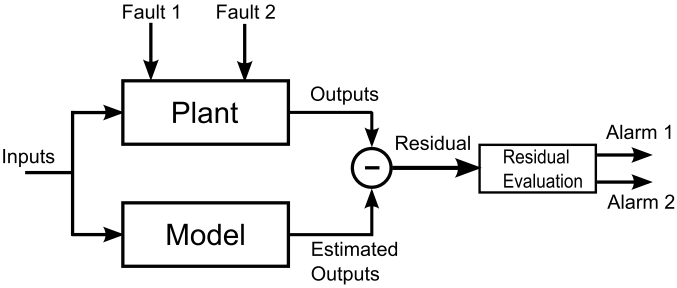


Figure 2.2: Block diagram of a generic model-based fault isolation scheme.

a given direction, the structured residuals approach where a bank of observers is jointly designed to ensure isolation. Two particular cases of the structured residuals approach are the Dedicated Observer scheme where each observer is sensitive to only one fault, and the Generalized Observer scheme where each observer is sensitive to all but one fault. The following example illustrates the latter method.

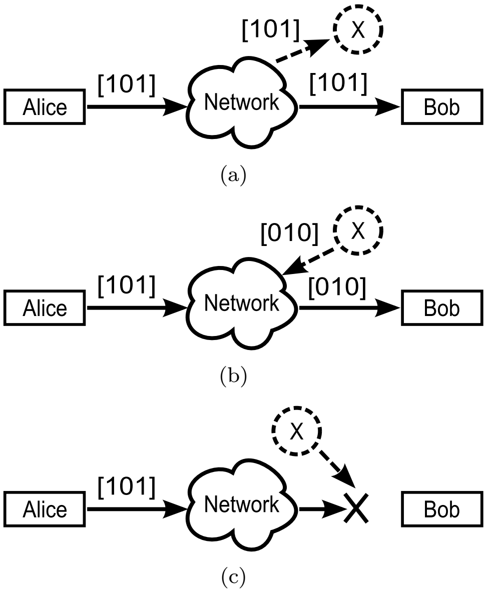
Example 2.3
Consider the static model in the previous examples with three faults

$$z = Hx + Ef + = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} x + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix},$$

where f_1, f_2, and f_3 are scalar faults. Assume only a single fault occurs. The Generalized Observer scheme is used to isolate the faults where three residuals are designed, each sensitive to all but one fault. Denoting r_1 as the residual insensitive to f_1, f_1 could be treated as a disturbance and r_1 generated using the approach in the Example 2.2. Repeating the procedure for r_2 and r_3, a bank of residuals is obtained with the following sensitivity table

	f_1	f_2	f_3
r_1	0	1	1
r_2	1	0	1
r_3	1	1	0

where 1 (0) denotes that the residual is sensitive to (decoupled from) a given fault. Given the former sensitivity table and the assumption that only a single fault occurs, the faults can be isolated once they are detected. The Generalized Observer scheme will be used in Chapter 5 for fault isolation in large-scale dynamical systems.



(a)

(b)

(c)

Figure 2.3: Examples with violation of (a) confidentiality, (b) integrity, and (c) availability.

2.2 IT Security

In this section we introduce the main concepts of IT security used in the thesis. Bishop (2002) identifies the fundamental properties of information and services in IT systems, namely

- confidentiality;
- integrity;
- availability.

Confidentiality concerns the concealment of data, ensuring it remains known to the authorized parties alone. Integrity relates to the trustworthiness of data, meaning there is no unauthorized change to the information between the source and destination. Availability considers the timely access to information or system functionalities. A possible violation of the system security is denoted as a threat, while an attack corresponds to the set of actions causing the violation to occur. The following examples are presented to illustrate the former concepts.

Example 2.4
Consider the examples illustrated in Figure 2.3. In all three cases, Alice is sending

the message `[101]` to Bob through a communication network. This is a private message, hence only Alice and Bob should know its contents.

In Figure 2.3(a), a third entity is able to eavesdrop on the communication, thus getting access to the message's contents. Therefore the confidentiality was violated.

Another scenario occurs in Figure 2.3(b), where the third entity succeeds in sending a wrong message to Bob, as if it was Alice sending it. Here data integrity is violated.

In our final example, illustrated in Figure 2.3(c), the message sent by Alice is actually blocked and does not reach Bob. Hence data availability was compromised.

The violations presented in the previous examples were caused by disclosure, deception, and Denial-of-Service attacks, respectively. Chapters 3 and 4 focus on deception attacks, similar to example in Figure 2.3(b)

Security Strategies

Building IT security then amounts to ensuring the confidentiality, integrity, and availability properties. There exist different approaches to handle cyber attacks, which can be divided among the following classes:

- prevention;
- detection;
- recovery.

Prevention takes place by removing or mitigating existing vulnerabilities, thus making attacks less likely. For instance, encryption of the communication link is a prevention method against disclosure attacks considered in Figure 2.3(a). In this case, unless the third party knows the encryption key, the message's contents cannot be accessed.

Detection, on the other hand, considers that attacks may take place. The objective is then to assess whether the system is under attack or not. In the deception attack considered in Figure 2.3(b), a possible detection method would be to analyze the contents of the message at destination, comparing them to what is expected. Such strategy is similar to model-based fault detection described in Section 2.1.1, and a model of the expected contents is required. These principles are indeed used in practice, for instance to detect abnormal data traffic using statistical models (Zhang et al., 2009). In our particular example, suppose Alice and Bob agreed that all the messages exchanged between them are required to have an even number of 1's. Since the message received by Bob has an odd number of 1's, Bob could then detect the data corruption.

Recovery deals with mitigation of the detected attacks, retaining a secure system state. In the case of the Denial-of-Service attack in Figure 2.3(c), one could have a mitigation scheme where the data is re-sent using a different path from source to destination, thus avoiding the compromised links.

Ensuring IT security is a dynamic procedure, as new vulnerabilities may be discovered by attackers, even within the prevention, detection, and recovery mechanisms. For instance, in the case of deception attacks the attacker may find attack strategies that bypass the current detection mechanisms, remaining undetected. Chapters 3 and 4 explore this particular scenario.

Remark 2.2.1. *The set of attacks bypassing given detection mechanisms are denoted as stealthy attacks. Recovering our example from Figure 2.3(b), if the fake message had been* `[110]` *it would bypass Bob's detection scheme, thus being a stealthy deception attack.*

2.3 Secure and Reliable Networked Control Systems

Secure NCS have received increasing attention recently. An overview of existing cyber threats and vulnerabilities in NCS is presented in (Cardenas et al., 2009; Cárdenas et al., 2008a,b), where the authors also mention some of the new challenges arising from the interconnection of IT and control systems. Particularly, realistic and rational adversary models are mentioned as one of the key items in security for NCS. For instance, physical attacks affect the physical plant directly and can thus be modeled as faults using the Fault Diagnosis literature mentioned in Section 2.1. In the framework of security, however, such faults are rational and are endowed with intelligence and intent. Therefore these faults may exploit vulnerabilities existing in the traditional fault detection mechanisms and remain undetected. In fact, Amin et al. (2010) reported experimental stealthy data deception attacks on water irrigation canals controlled by SCADA systems. Smith (2011) characterized stealthy attack policies for scenarios where the attacker is able to perform disclosure and deception attacks on all the sensors, illustrating it on the same water irrigation system.

Rational attackers performing stealthy deception attacks were also considered for sensor networks distributively computing linear functions, where each node is modeled as a first-order system (Pasqualetti et al., 2011; Sundaram and Hadjicostis, 2011; Sundaram et al., 2010). The class of stealthy deception attacks was characterized in a system-theoretic fashion in terms of the number of compromised nodes and the network connectivity. Other work also considering rational attackers analyzes Denial-of-service attacks, where the optimal attack policy under finite resources is characterized (Amin et al., 2009; Gupta et al., 2010).

Mo and Sinopoli (2009) considered replay attacks on wireless networks performing state estimation, which are a particular class of deceptions attacks, and proposed a novel detection scheme tailored to this class of attacks. The safety of Automatic Generation Control for power system under deception attacks was considered in (Esfahani et al., 2010) and the authors showed that the cyber attacks could violate the system safety constraints. Although the attackers were not assumed to be rational in these papers, the former illustrates that tailored detection mechanisms can increase the attack detection rate, while the latter identified existing safety vulnerabilities.

Benchmark examples for NCS security were described in (Rieger, 2010) and numerical experiments on a benchmark process plant were reported in (Cárdenas et al., 2011). In the latter, although a mathematical formulation for the effects of cyber attacks was given, as well as attack objectives, the attack policies presented did not make use of the full attacker resources.

Security of NCS is a recent research field, full of broad open questions. Furthermore recent work within the control community, this thesis included, only consider how the possible threats affect the control system dynamics, while the IT system and respective security mechanisms are neglected. Although this approach is interesting and valid, looking more closely at the IT framework and tools may help to design better protective schemes.

2.4 Power Transmission Networks

SCADA systems in power transmission networks have evolved substantially since they were introduced in the 1960s (Wu et al., 2005). The early systems were mainly used for logging data from the power network. Today modern SCADA systems are enhanced by Energy Management Systems (EMS) providing system-wide monitoring and control to meet performance and reliability constraints (Balu et al., 1992; Shahidehpour et al., 2005a).

Due to constraints of traditional technologies, only quasi-steady state dynamics are captured by current SCADA EMS. However, with the advent of new sensors such as Phasor Measurement Units (PMUs), slow transient behaviors of power transmission networks can now be captured. This leads to the so-called Wide-Area Monitoring and Control Systems (WAMS/WAMC), providing yet another layer of control with increased performance and reliability. These systems are briefly described in the remaining of this section.

2.4.1 Energy Management System

Figure 2.4 illustrates some of the components in traditional SCADA EMS systems. Power networks are hybrid systems, having analog variables, such as voltages and currents, and digital variables like breaker status. System-wide measurements of these variables are taken locally at the substation level, gathered by RTUs, and transmitted to the control center through the communication network. Since not all variables are measured, the current state of the power network needs to be estimated based on the received measurements and a detailed system model. The optimal state and measurement estimates are computed by the state estimator (SE), see Figure 2.4. Possible measurement errors biasing the estimates can be handled *a posteriori* by bad data detectors (BDD) using model-based approaches as mentioned in Section 2.1.1.

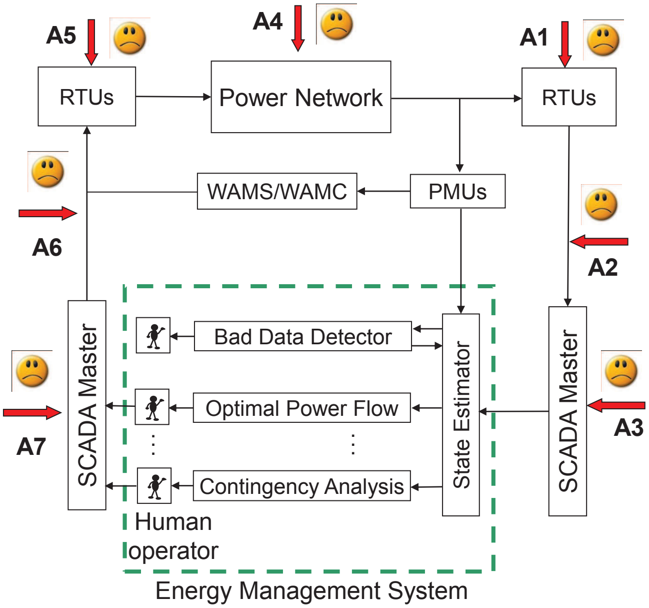


Figure 2.4: A schematic block diagram of a power network, a SCADA system, a control center, and possible IT vulnerabilities.

The SE provides system observability to operators and other EMS tools, thus being an integral tool in power network operation. As shown in Figure 2.4, contingency analysis (CA) tools use the estimates to evaluate if the system meets the required reliability criteria in the presence of hypothetical equipment failure. On the other hand, optimal power flow (OPF) analysis based on the estimates evaluates possible improvements in performance. Based on the recommendations from the CA and OPF, the human operator chooses suitable control actions to be applied to the power network, as illustrated in Figure 2.4.

Cyber threats

There are several threats in a SCADA system, given the complexity of the IT system and the large number of heterogeneous components and services therein. In Figure 2.4 we illustrate some of these threats and the respective entry points to the SCADA EMS. For instance, the entry points A1–A3 and A5–A7 for the different attacks discussed in the previous section. The measurements sent by the RTU (A2) and the system information in the SCADA databases (A3 and A7) could be targets of disclosure attacks to gain access to confidential data, such as the power network model. A Denial-of-Service attack could be performed on the communication links between the RTUs and the control center (A2 and A6), resulting in loss of avail-

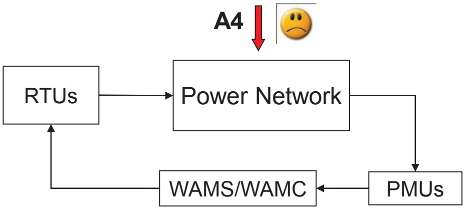


Figure 2.5: A schematic block diagram of a power network with a WAMS monitoring faults.

ability. Another attack scenario corresponds to deception attacks on the RTU data sent to the control center (A1–A3), resulting in a violation of data integrity. This scenario is further discussed in Chapters 3 and 4, where the we characterize the class of stealthy deception attacks bypassing existing detection schemes, similar to the scenario illustrated in Figure 2.3(b) and discussed in the Example 2.4.

2.4.2 Wide-Area Monitoring and Control Systems

Power transmission networks are large-scale spatially distributed systems operating under strict safety and reliability constraints (Shahidehpour et al., 2005b). Monitoring the overall state of the system is essential to ensure awareness of the system operating conditions and timely detect events that may disrupt the power network operation.

As described earlier, currently monitoring schemes are implemented in a centralized control center through a single state estimator. The core methodology for state estimation of power systems dates from 1970 (Abur and Exposito, 2004; Schweppe and Wildes, 1970). Due to the low sampling frequency of the sensors in these systems a steady-state approach is taken and reliability is ensured by overconstraining the network operation. Furthermore dynamic faults such as generator electro-mechanical oscillations may pass undetected by schemes based on steady-state models and measurements, possibly leading to cascade failures.

Recently measurement units with higher sampling rate such as the PMUs have been developed, opening the way to dynamic state estimators and model-based fault detection schemes taking into account the dynamics of the system. An example of the new opportunities is the Wide-Area Monitoring System (WAMS), which uses data from several PMUs to perform real-time monitoring of large-scale power transmission networks (Machowski et al., 2008). Several implementations of WAMS have recently been performed, which have proven useful in monitoring and damping power system oscillations (Phadke and de Moraes, 2008). In a recent survey, Chompoobutrgool et al. (2011) present an overview of possible uses for WAMS, such as dynamic state estimation and fault monitoring through Kalman filters.

In this thesis we consider distributed monitoring of faults in the power transmission network.

These technological developments allow for new opportunities to be envisioned, such as a PMU-enabled WAMS monitoring for the system for physical faults illustrated in Figure 2.5. This serves as motivation for the contributions in Chapter 5, where a distributed model-based fault monitoring scheme is proposed.

Chapter 3

Cyber-Security of State Estimator in Power Systems

This chapter considers the cyber security of SCADA EMS providing support to human operators for reliable and optimized power network operation, as described in Section 2.4. Specifically, we consider that the IT system integrity is violated due to deception attacks on the measurement data, as discussed in Section 2.2. The current detection mechanisms for measurement errors (BDD) are analyzed to identify existing vulnerabilities, leading to the characterization of stealthy deception attacks that are able to remain undetected.

The outline of this chapter is as follows. Section 3.1 gives an overview of related work and summarizes this chapter's contributions. The theoretical concepts behind the SE and BDD algorithms in power networks are presented in Section 3.2 and the VIKING 40$-$bus benchmark used in the numerical examples is described in Section 3.3. In Section 3.4 a novel attacker model and optimization framework to analyze the vulnerability to stealthy attacks is proposed. Some considerations regarding limitations of linear attack policies are also given and a security metric quantifying each measurement's vulnerability to stealthy attacks is proposed. Section 3.5 contains the analysis of stealthy attack based on perturbed models and the description and results of experiments conducted in a SCADA EMS software. A summary of the chapter is presented in Section 3.6.

3.1 Contributions and Related Work

In current implementations of SE algorithms, there are bad data detection (BDD) schemes designed to detect random outliers in the measurement data based on a detailed measurement model and high measurement redundancy. However, these methods may fail in the presence of an intelligent attacker. For instance, it is well known (Abur and Exposito, 2004; Monticelli, 1999) that the BDD can fail to detect and isolate the faulty measurements for the so-called *multiple interacting bad data*, which are correlated errors affecting several measurements usually caused by

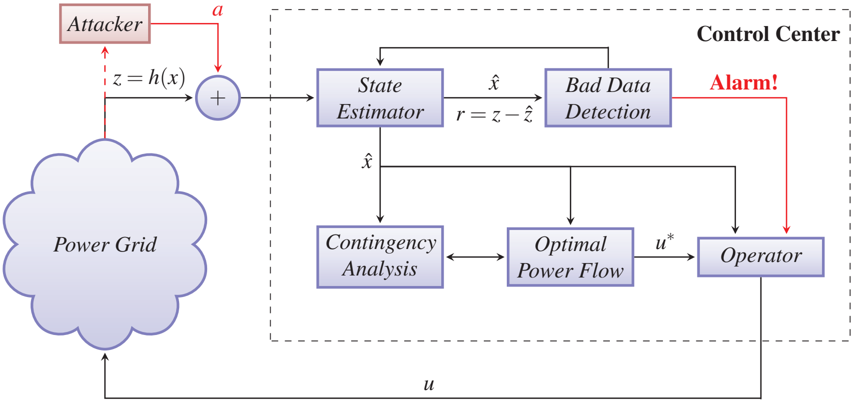


Figure 3.1: The state estimator under a cyber attack

parameter and topology model mismatches.

Consider Figure 3.1 illustrating the SCADA EMS in power networks under deception attacks on the measurement data. As described in Section 2.4, system-wide measurements of physical variables are taken and gathered locally at the substations by the RTUs. The total set of measurements arranged in a vector is denoted as $z \in \mathbb{R}^m$ in Figure 3.1. The deception attack on the measurements is modeled by $a \in \mathbb{R}^m$, where each element a_i corresponds to the data corruption added to the corresponding measurement, z_i. The vector of corrupted measurement data used by the EMS components is then $z^a = z + a$.

A class of data corruption attacks a undetectable by conventional BDD schemes has been characterized in recent work (Bobba et al., 2010; Giani et al., 2011; Liu et al., 2009; Sandberg et al., 2010). Several countermeasures to these attacks were proposed, from the allocation of encryption (Dán and Sandberg, 2010) and additional protected measuring devices (Giani et al., 2011), to the implementation of improved BDD schemes, see (Bobba et al., 2010; Kosut et al., 2010). Methods to efficiently rank the measurements in terms of their vulnerability and finding sparse attacks requiring the corruption of a low number of measurements were also proposed (Dán and Sandberg, 2010; Giani et al., 2011; Sandberg et al., 2010; Sou et al., 2011). However, the aforementioned class of undetectable attacks was derived assuming that both the BDD and the attacker know exactly a simplified linear network model. This assumption on the attacker knowledge may be too restrictive in most scenarios. Moreover, the real power network is nonlinear and a nonlinear model is also typically implemented in the SE. Therefore, it is not clear how a nonlinear SE will react to these stealthy deception attacks or how large a can be before the SE no longer converges.

Two main contributions regarding stealthy attacks on SCADA EMS are reported in this chapter. First, a novel attacker model is proposed, including attack goals and constraints in addition to being stealthy. Given the characterization of stealthy attacks using accurate linear models, the proposed attacker model directly leads to attack policies formulated in an optimization framework. The proposed framework may be used to analyze each measurement's vulnerability to stealthy attacks (Sandberg et al., 2010) and suggest defensive solutions (Dán and Sandberg, 2010).

As our second contribution we relax the assumption that the attacker has an accurate linear model of the system and consider perturbed models, for instance due to linearization and varying operating conditions. In this framework we provide bounds on the attack vector size ensuring that the attack is stealthy, despite the model errors. Then we consider a nonlinear SE algorithm and present experimental results on its sensitivity to stealthy deception attacks computed using linear policies. To the best of our knowledge, this is the first such experimental study on nonlinear SE algorithms. Maybe somewhat surprisingly, for the cases we have studied, the attacks indeed pass undetected for very large corruptions a, indicating that the simplifying assumptions used in the previously mentioned studies are indeed valid.

These contributions increase the understanding regarding vulnerabilities in current detection mechanisms, which is an essential step in improving the system's security and reliability.

3.2 Preliminaries

In this section we introduce the power network models and the theory behind the SE, BDD, and OPF algorithms.

3.2.1 Measurement Model

For an $N-$bus electric power network, the $n = 2N - 1$ dimensional state vector x is $(\theta^\top, V^\top)^\top$, where $V = (V_1, \dots, V_N)$ is the vector of bus voltage magnitudes and $\theta = (\theta_2, \dots, \theta_N)$ vector of phase angles. This state vector is the minimal information needed to characterize the operating point of the power network. Without loss of generality, we have considered bus 1 to be the reference bus, hence all phase-angles are taken relatively to this bus and $\theta_1 = 0$. The $m-$dimensional measurement vector z can be grouped into two categories: (1) z_P, the active power flow measurements P_{ij} from bus i to j and active power injection measurement P_i at bus i, and (2) z_Q, the reactive power flow measurements Q_{ij} from bus i to j, reactive power injection measurement Q_i and V_i voltage magnitude measurement at bus i. The neighborhood set of bus i, which consists of all buses directly connected to this bus, is denoted by N_i. The power injections at bus i are described by

$$\begin{array}{rcl} P_i & = & V_i \sum_{j \in N_i} V_j \left(G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij} \right) \\ Q_i & = & V_i \sum_{j \in N_i} V_j \left(G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij} \right) \end{array},$$

and the power flows from bus i to bus j are described by

$$\begin{array}{rcl} P_{ij} & = & V_i^2(g_{si}+g_{ij}) - V_iV_j\left(g_{ij}\cos\theta_{ij} + b_{ij}\sin\theta_{ij}\right) \\ Q_{ij} & = & -V_i^2(b_{si}+b_{ij}) - V_iV_j\left(g_{ij}\sin\theta_{ij} - b_{ij}\cos\theta_{ij}\right) \end{array},$$

where $\theta_{ij} = \theta_i - \theta_j$ is the phase angle difference between bus i and j, g_{si} and b_{si} are the shunt conductance and susceptance of bus i, g_{ij} and b_{ij} are the conductance and susceptance of the branch from bus i to j, and $Y_{ij} = G_{ij} + jB_{ij}$ is the $(i,j)-$th entry of the nodal admittance matrix. More detailed formulas relating measurements z and state x may be found in (Abur and Exposito, 2004).

Assuming that the model parameters and the network topology are exact, the nonlinear measurement model for state estimation is defined by

$$z = h(x) + \epsilon, \tag{3.1}$$

where $h(\cdot)$ is the $m-$dimensional nonlinear measurement function that relates measurements to states and is assumed to be twice continuously differentiable, $\epsilon = (\epsilon_1, \ldots, \epsilon_m)^\top$ the zero mean measurement error vector, and usually $m \gg n$ meaning that there is high measurement redundancy. Here ϵ_i are independent Gaussian variables with respective variances σ_i^2 indicating the relative uncertainty about the $i-$th measurement and thus we have $\epsilon \sim \mathcal{N}(0, R)$ where $R = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$ is the covariance matrix.

DC measurement model

The so-called DC network model is a linear measurement model obtained by neglecting the reactive power components and assuming that the voltage magnitudes are constant at 1pu (per unit), there are no branch resistances and shunt admittances, and the phase-angles θ_i are close to zero, leading to the following measurement equations

$$\begin{array}{c} P_i = \sum_{j \in N_i} b_{ij}(\theta_i - \theta_j) \\ P_{ij} = -b_{ij}(\theta_i - \theta_j) \end{array}.$$

The resulting linear measurement model is then given by

$$z = H_{DC}x + \epsilon, \tag{3.2}$$

where z contains only active power measurements and the state x corresponds to the phase-angles, i.e. $x = \theta$.

3.2.2 State Estimator

The basic SE problem is to find the best n-dimensional state x for the measurement model (3.1) in a weighted least square (WLS) sense. Defining the residual vector

$r(x) = z - h(x)$, we can write the WLS problem as

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} J(x) = \frac{1}{2} r(x)^\top R^{-1} r(x) \\ &\text{such that} \quad g(x) = 0, \quad s(x) \leq 0, \end{aligned} \tag{3.3}$$

where the inequality constraints generally model saturation limits, while the equality constraints are used to include target setpoints and to ensure physical laws such as zero power injection transition buses, e.g., transformers, and zero power flow in disconnected branches. Such data used in the equality constraints is often seen as *pseudo-measurements*. For sake of simplicity, we will present the solution to the unconstrained optimization problem.

The unconstrained WLS problem is posed as

$$\min_{x\in\mathbb{R}^n} J(x) = \frac{1}{2} r(x)^\top R^{-1} r(x).$$

The SE yields a *state estimate* $\hat{x}$ as a minimizer to this problem. The solution $\hat{x}$ can be found using the *Gauss-Newton method* which solves the so called *normal equations*:

$$\left(H^\top(x^k) R^{-1} H(x^k)\right) \Delta x^k = H^\top(x^k) R^{-1} r(x^k), \tag{3.4}$$

for $k = 0, 1, \ldots$, where

$$H(x^k) := \left.\frac{dh(x)}{dx}\right|_{x=x^k}$$

is called the Jacobian matrix of the measurement model $h(x)$ and $\Delta x^k = x^{k+1} - x^k$. The normal equations yield a unique solution if the measurement Jacobian matrix $H(x^k)$ is full column rank. In this case, the power network is said to be observable. Consequently, the matrix $\left(H^\top(x^k) R^{-1} H(x^k)\right)$ in (3.4) is positive definite and the Gauss-Newton step generates a descent direction, i.e., for the direction $\Delta x^k = x^{k+1} - x^k$ the condition $\nabla J(x^k)^\top \Delta x^k < 0$ is satisfied. The estimation algorithm is finalized when the stopping criteria

$$\Delta x^k = \left(H^\top(x^k) R^{-1} H(x^k)\right)^{-1} H^\top(x^k) R^{-1} r(x^k) \approx 0$$

is met and the optimal estimate is taken as $\hat{x} = x^k$.

For notational convenience, throughout the next sections we will use $H(x^k)$ as H, Δx^k as Δx, and $r(x^k) = z - h(x^k)$ as r.

DC state estimator

Considering the linear DC model (3.2), the SE problem (3.3) reduces to a constrained linear least squares problem. In the unconstrained case the optimal estimate is obtained using the normal equations

$$\hat{x} = (H_{DC}^\top R^{-1} H_{DC})^{-1} H_{DC}^\top R^{-1} z.$$

Remark 3.2.1. *Henceforth we consider the covariance matrix R to be the identity matrix, for simplicity, i.e., all measurements have unitary weights.*

3.2.3 Bad Data Detection

Through BDD the SE detects measurements corrupted by errors whose statistical properties exceed the presumed standard deviation or mean (Abur and Exposito, 2004). As mentioned in Section 2.1.1, this can be achieved by hypothesis tests using the statistical properties of the measurement residual (3.5), which is now characterized under the presence of measurement errors.

Proposition 3.2.1. *Consider the measurements $z = h(x) + \epsilon$ and suppose the optimal estimate in the least squares sense, $\hat{x}$, is obtained with the Gauss-Newton method. The first-order approximation of the measurement residual $r(\hat{x}) = z - h(\hat{x})$ is given by*

$$r = S\epsilon, \tag{3.5}$$

where $S = I - H(H^\top H)^{-1}H^\top$ and $H = \left.\frac{\partial h(x)}{\partial x}\right|_{x=\hat{x}}$.

Proof. Performing a first-order approximation on the measurement model, $h(x) \approx h(\hat{x}) + H(x - \hat{x})$, and applying it to the measurement residual we obtain $r \approx h(\hat{x}) + H(x - \hat{x}) + \epsilon - h(\hat{x}) = H(x - \hat{x}) + \epsilon$. Since the Gauss-Newton method has converged, the stopping criteria $(H^\top H)^{-1}H^\top r(\hat{x}) \approx 0$ is met and, together with the residual first-order approximation, results in $H(x - \hat{x}) = -H(H^\top H)^{-1}H^\top \epsilon$. The residual first-order approximation is then given by $r = (I - H(H^\top H)^{-1}H^\top)\epsilon$. □

Remark 3.2.2. *The measurement estimate (residual) sensitivity matrix corresponds to the orthogonal projector onto* $\mathrm{Im}(H)$ *(*$\mathrm{Ker}(H^\top)$*), defined as $\mathbf{P}_{\mathrm{Im}(H)} = H(H^\top H)^{-1}H^\top$ ($\mathbf{P}_{\mathrm{Ker}(H^\top)} = I - H(H^\top H)^{-1}H^\top$). Some properties of orthogonal projectors will be useful in the following sections, namely $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^\top$.*

Remark 3.2.3. *An expression similar to* (3.5) *can be obtained for the measurement residual in the DC-SE by replacing H with H_{DC}.*

We now introduce two of the BDD hypothesis tests widely used in practice, the *performance index test* and the *largest normalized residual test*, which will be used when characterizing stealthy attack policies.

Performance index test

Consider the quadratic cost function evaluated at the optimal estimate $\hat{x}$

$$J(\hat{x}) = r^\top r = \epsilon^\top S\epsilon. \tag{3.6}$$

Recalling that $\mathrm{rank}(H) = n$, $\mathrm{Im}(H) \oplus \mathrm{Ker}(H^\top) = \mathbb{R}^m$, and $S = \mathbf{P}_{\mathrm{Ker}(H^\top)}$, from the definition of orthogonal projector we have $\mathrm{rank}(S) = m - n$. Therefore, in the absence of bad data, the quadratic form $\epsilon^\top S\epsilon$ has a chi-squares distribution with $m-n$ degrees of freedom, *i.e.* $J(\hat{x}) \sim \chi^2_{m-n}$ with $\mathbb{E}\{J(\hat{x})\} = m-n$. The main idea behind the performance index test is to use $J(\hat{x})$ as an approximation of y and check if $J(\hat{x})$ follows the distribution χ^2_{m-n}. This can be posed as a hypothesis test with a null hypothesis H_0, which if accepted means there is no bad data, and an alternative bad data hypothesis H_1 where

$$H_0 : \mathbb{E}\{J(\hat{x})\} = m-n, \quad H_1 : \mathbb{E}\{J(\hat{x})\} > m-n$$

Defining $\alpha \in [0,\ 1]$ as the significance level of the test corresponding to the false-alarm rate, and $\tau_\chi(\alpha)$ such that

$$\int_0^{\tau_\chi(\alpha)} g^\chi(u)du = 1-\alpha, \tag{3.7}$$

where $g^\chi(u)$ is the probability distribution function (pdf) of χ^2_{m-n}, and noting that $J(\hat{x}) = \|r(\hat{x})\|_2^2$ the result of the test is

$$\begin{aligned} \text{reject } H_0 \text{ if } \|r\|_2 &> \sqrt{\tau_\chi(\alpha)}, \\ \text{accept } H_0 \text{ if } \|r\|_2 &\leq \sqrt{\tau_\chi(\alpha)}. \end{aligned}$$

Largest normalized residual test

Recall that from (3.5) we have $r \sim \mathcal{N}(0,S)$ and consider the weighted residual vector

$$r^N = D^{-1/2}r, \tag{3.8}$$

with $D \in \mathbb{R}^{m\times m}$ being a diagonal matrix defined as $D = \text{diag}(S)$. In the absence of bad data each element r_i^N, $i = 1,\dots,m$ of the weighted residual vector then follows a normal distribution with zero mean and unit variance, *i.e.* $r_i^N \sim \mathcal{N}(0,1)$, $\forall i = 1,\dots,m$, and r^N is often named as the *normalized residual.* Hence bad data could be detected by checking if r_i^N follows $\mathcal{N}(0,1)$, which may be posed as the following hypothesis test:

$$H_0 : \mathbb{E}\left\{r_i^N\right\} = 0, \quad H_1 : \mathbb{E}\left\{|r_i^N|)\right\} > 0.$$

Again defining $\alpha \in [0,\ 1]$ as the significance level of the test and $\tau_\mathcal{N}$ such that

$$\int_{-\tau_\mathcal{N}(\alpha)}^{\tau_\mathcal{N}(\alpha)} g^\mathcal{N}(u)du = 1-\alpha, \tag{3.9}$$

where $g^\mathcal{N}(u)$ is the pdf of $\mathcal{N}(0,1)$, and noting (3.8), the result of the test is

$$\begin{aligned} \text{reject } H_0 \text{ if } \|D^{-1/2}r\|_\infty &> \tau_\mathcal{N}(\alpha) \\ \text{accept } H_0 \text{ if } \|D^{-1/2}r\|_\infty &\leq \tau_\mathcal{N}(\alpha) \end{aligned}$$

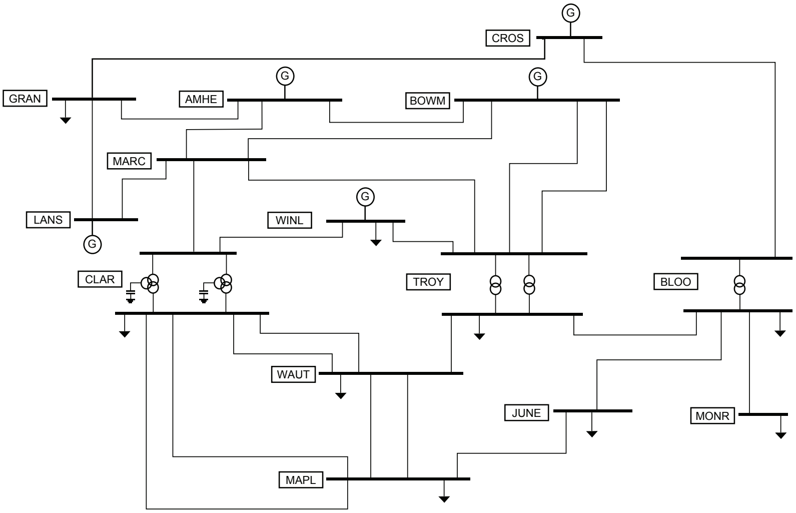


Figure 3.2: Power network considered in the experiment.

Remark 3.2.4. *Denote* $\| \cdot \|_p$ *as the* $p-$*norm for* $p \geq 0$ *and consider* $\| \cdot \|_0$ *to be the cardinality of the respective vector. Clearly both tests may be written as*

$$\|Wr(\hat{x})\|_p \underset{H_1}{\overset{H_0}{\lessgtr}} \tau, \tag{3.10}$$

for suitable weights W, *norms* p, *and thresholds* τ.

3.3 VIKING 40−bus Benchmark

In this thesis the VIKING 40−bus power transmission network is used as a benchmark example to illustrate some of our contributions through numerical simulations and experiments. This benchmark was designed under the EU FP7 VIKING project (Björkman, 2010; Giani et al., 2009). This network, similar to the IEEE 39−bus network, is shown in Figure 3.2. The system consists of 14 substations and the bus-branch model has 27 buses and 40 branches. The default measurement configuration, i.e., which physical variables are measured, renders the power network observable but does not include all the possible measurements.

The power network topology, model parameters, and measurement configuration were imported to MATLAB using the MATPOWER toolbox (Zimmerman et al., 2009), mirroring the network model information available in a SCADA EMS system. The network model and EMS component algorithms from the VIKING 40−bus benchmark will be used to illustrate the framework developed in this chapter and also in Chapter 4.

3.4 State Estimation under Stealthy Deception Attacks

Using the theory and models described in the previous section, we present the framework used throughout the next sections to study the cyber security of SCADA EMS software and algorithms.

3.4.1 Attacker Model

Let the corrupted measurement be denoted z^a. We assume the following additive attack model

$$z^a = z + a, \tag{3.11}$$

where $a \in \mathbb{R}^m$ is the attack vector introduced by the attacker, see also Figure 3.1. The vector a has zero entries for uncompromised measurements.

The purpose of a stealthy deception attacker is to compromise the telemetered measurements available to the SE while meeting the following conditions:

1. the SE algorithm converges;

   Since the SE convergence is not in the scope of this thesis, this condition is assumed to be met. Alternatively, one could constrain the attack to be sufficiently small so that convergence is indeed achieved. Later in Section 3.5.3 an experimental example is shown where this condition was violated for large attacks.

2. the attack remains undetected by the BDD scheme;

   The attacker's action will be undetected by the BDD scheme provided that the measurement residual under attack, $r^a := r(\hat{x}^a) = z^a - h(\hat{x}^a)$, satisfies the condition (3.10) for H_0, i.e., $\|Wr(\hat{x})\|_p < \tau$. We will occasionally use the notation $\hat{x}^a(z^a)$ to emphasize the dependence on z^a.

3. for the targeted set of measurements, the estimated values at convergence are close to the compromised ones introduced by the attacker;

   Consider that the attacker aims at corrupting measurement i. This means the attacker would like the estimated measurement $\hat{z}_i^a := h_i(\hat{x}^a(z^a))$ to be equal to the actual corrupted measurement z_i^a. Therefore, we consider that the attack vector a is chosen such that $|z_i^a - \hat{z}_i^a| = 0$.

The aim of a stealthy deception attacker is then to find and apply an attack a that satisfies conditions 1), 2), and 3). This problem can be posed as

$$\begin{aligned} &\text{find } a \\ &\text{s.t. } \; a \in \mathcal{G} \cap \mathcal{C} \cap \mathcal{U} \;, \end{aligned} \tag{3.12}$$

where $\mathcal{G}$ is the set of goals in condition 3), $\mathcal{C}$ the set of constraints ensuring condition 1) is met and that no protected measurements are corrupted, and $\mathcal{U}$ the set of stealthy attacks satisfying condition 2).

Stealthy attacks based on linear models

In general a stealthy attack requires the corruption of more measurements than the targeted one, as a stealthy attack must have the attack vector a fitting the measurement model in order to bypass the BDD (Liu et al., 2009; Mili et al., 1985). This result has been illustrated for the DC-SE (Liu et al., 2009), where the class of stealthy attacks for linear DC model is characterized by $a \in \mathrm{Im}(H_{DC})$. The following statement extends the previous concept to the nonlinear SE, which has been discussed by Mili et al. (1985) regarding multiple interacting bad data.

Theorem 3.4.1. *Consider the measurement residual under attack $r^a = z^a - h(\hat{x}^a)$ with $z^a = z + a$, $z = h(x) + \epsilon$, and $\epsilon \sim \mathcal{N}(0, I)$ and let $H = \left.\frac{\partial h(x)}{\partial x}\right|_{x=\hat{x}^a}$. For optimal estimates obtained by the Gauss-Newton method, attacks satisfying $a \in \mathrm{Im}(H)$ are not detected by any BDD test of the form $\|Wr^a\|_p \underset{H_1}{\overset{H_0}{\lessgtr}} \tau$.*

Proof. From (3.5) we obtain the following expression for the residual under attack $r^a = S(a + \epsilon)$. Recalling that $\mathrm{Ker}(H^\top)$ and $\mathrm{Im}(H)$ are orthogonal subspaces and $S = \mathbf{P}_{\mathrm{Ker}(H^\top)}$, it immediately follows that for any $a \in \mathrm{Im}(H)$ we have $r^a = S(a + \epsilon) = S\epsilon$, which meets the statistical models under no bad data, H_0, as discussed in Section 3.2.3. □

Remark 3.4.1. *Note that computing these attacks requires knowledge of the linearized model at the biased estimate $\hat{x}^a$, $H = \left.\frac{\partial h(x)}{\partial x}\right|_{x=\hat{x}^a}$, which is not known a* priori.

Therefore, given the above remark, it remains unclear whether attacks computed using linear models $H(x) \neq H(\hat{x}^a)$ are indeed stealthy for nonlinear estimators. This issue is further analyzed in Section 3.5.

In the remainder of this section, the attack vector a is assumed to be sufficiently small such that $\mathrm{Im}\left(H(x)\right) \approx \mathrm{Im}\left(H(\hat{x}^a)\right)$, implying that $S(x) \approx S(\hat{x}^a)$. Therefore non-trivial attacks of the form

$$a = Hc, \quad c \notin \mathrm{Ker}(H) \tag{3.13}$$

are assumed undetected by the BDD. Following this approach we can provide more insights on the feasibility and structure of the attacks, based on which protective schemes can be designed, as discussed in the following sections.

Remark 3.4.2. *The class of stealthy attacks for the DC-SE, $a \in \mathrm{Im}(H_{DC})$, immediately follows by replacing H with H_{DC}.*

3.4.2 Minimum Effort Attack Synthesis

We now present a general methodology for synthesizing stealthy attacks based on linear models with specific target constraints. Suppose the attacker wishes to compute the "least-effort" attack in the p-norm sense satisfying (3.12). Assuming the attacker knows the linear model H, such attack could be computed by solving the optimization problem

$$\begin{aligned} &\min_a \|a\|_p \\ &\text{s.t. } a \in \mathcal{G} \cap \mathcal{C} \cap \mathcal{U} \ , \end{aligned} \tag{3.14}$$

where $\mathcal{U} = \operatorname{Im}(H)$. A particular formulation is the 2-norm case with a single attack target, $\mathcal{G} = \{a \in \mathbb{R}^m : a_k = 1\}$. The objective is then to introduce a bias of 1 in measurement k, while minimizing $\|a\|_2$. Given (3.13), the optimization problem may be recast as

$$\begin{aligned} &\min_c \|Hc\|_2^2 \\ &\text{s.t. } e_k^\top Hc = 1 \end{aligned} , \tag{3.15}$$

where e_k is a unitary vector with 1 in the i-th component. Recall $K = \mathbf{P}_{\operatorname{Im}(H)} = HH^\dagger$.

Proposition 3.4.2. *The solution* a^* *to the optimization problem* (3.15) *is given by* $a^* = \frac{K}{K_{kk}} e_k$

Proof. The Lagrangian of this optimization problem is $L(c,\nu) = cH^\top Hc + \nu(e_k^\top Hc - 1)$ and the KKT conditions (Boyd and Vandenberghe, 2004) for an optimal solution (c^*, ν^*) are

$$\begin{cases} H^\top Hc^* + \nu^* H^\top e_k = 0 \\ \qquad\quad e_k^\top Hc^* - 1 = 0 \end{cases} . \tag{3.16}$$

Assuming the power network is observable, the solution for the first equation is $c^* = \nu^* H^\dagger e_k$. Including this in the second equation results in $\nu^* e_k^\top K e_k = 1$ which is equivalent to $\nu^* = \frac{1}{K_{kk}}$ with K_{kk} being the i-th diagonal element of K. We then have that $a^* = Hc^* = \frac{K}{K_{kk}} e_k$. □

In the power system's literature, the hat matrix K is known to have information regarding measurement redundancy and correlation. This result highlights a new meaning: each column of K actually corresponds to a $2-$norm optimal attack vector yielding a zero residual.

Another interesting case is that of $p = 0$, which means the attacker is computing the attack with minimum cardinality, e.g., minimizing the number of sensors to corrupt. A more detailed discussion of this case follows next.

3.4.3 A Security Metric

In Sandberg et al. (2010) the vulnerability of each measurement k was evaluated by studying the following problem:

Problem 1. *Given a data attack targeting measurement k, what is the minimal number of attacked measurements so that the data attack is undetectable by the Bad Data Detection?*

This problem was formulated as an optimization problem

$$\begin{aligned} \rho_k &= \min_{c} \|a\|_0 \\ a &= Hc \qquad , \\ 1 &= e_k^\top a \end{aligned} \tag{3.17}$$

where e_k is a vector of zeros with the $k-$th entry set to 1. The resulting optimal value ρ_k is then taken as a security index for measurement k. Defensive actions to secure the state estimator using this security index for the DC-SE were discussed in (Dán and Sandberg, 2010), while efficient algorithms to solve (3.17) have been proposed by Sou et al. (2011).

Numerical example on the VIKING $40-$bus benchmark

As described in previous sections, some information about the power network is needed to compute stealthy deception attacks. Here we consider a particular class of such information, namely the bus-branch model of the network. In this experiment we use the VIKING $40-$bus benchmark described in Section 3.3, where the model information was imported to MATLAB. Instead of the nonlinear model, only the DC model of the network is used, namely H_{DC} described in Section 3.2. Recall that in this model only active power and phase-angles are considered.

The algorithm in (Dán and Sandberg, 2010) was used to compute the security metrics for each measurement. Information regarding which measurements were assumed to be tamper-proof was taken into account, i.e., pseudo-measurements resulting from physical laws such as a transformer having zero power injection.

The result is presented in Figure 3.3. Given the default measurement configuration of the VIKING $40-$bus benchmark, we computed the security metric ρ_k (the red full circles) as defined in Section 3.4.1, obtaining quite heterogeneous results. Recalling that ρ_k is the minimum number of measurements needed to perform a stealthy attack on measurement k, we conclude that measurements with low ρ_k are relatively easily attacked while the ones with $\rho_k = \infty$ are fully protected.

Increasing the redundancy of the system by adding more measurements to the SCADA system increases the security level, as we see by looking at how $\bar{\rho}_k$ is larger than ρ_k, since $\bar{\rho}_k$ is the security metric computed assuming that all possible measurements are being taken. However, note that this does not guarantee full protection, as all measurements with finite ρ_k still have finite $\bar{\rho}_k$.

3.4.4 Limitations of Linear Policies

Recall from Theorem 3.4.1 that the core of the linear policies for stealthy attacks is to have $a \in \operatorname{Im}(H)$, where H depends on the corrupted estimate under the attack,

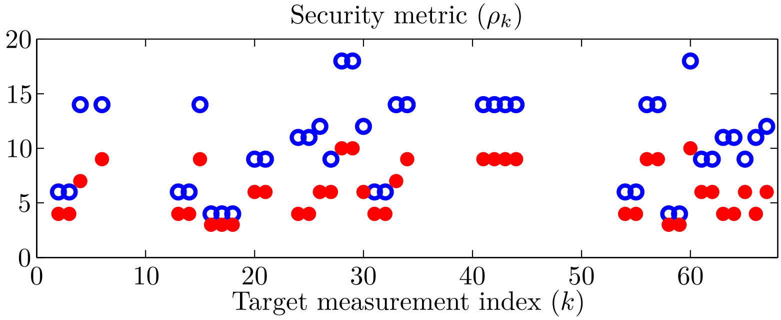


Figure 3.3: Security metrics for each measurement k: ρ_k (red full circles) was computed considering the default measurement configuration, while $\bar{\rho}_k$ (blue rings) was computed assuming that all possible measurements are taken. Both represent the minimum number of measurements needed to stealthily attack the target measurement k.

therefore unknown *a priori*. Even if the attacker is able to obtain a linearized model for other operating conditions, it is yet unclear if the resulting attacks would be stealthy. Furthermore, there are other constraints that were not considered in the previous discussion, such as saturation. These issues represent limitations on the linear attack policies described in Section 3.4.1 and are now briefly discussed.

Saturation limits.

Since the linear stealthy attack policy (3.13) is obtained by linearizing the nonlinear model (3.1) at a given state x, this policy is only valid in a region close to the true state x. This fact is particularly important when the saturation occurs. From considering (3.13) alone we do not have any limits on the size of $a = Hc$. However, the nonlinear model clearly shows that the measurements have saturation limits. For instance, disregarding the line and shunt conductances g_{ij} and g_{si} we have $P_{ij} = -V_i V_j b_{ij} \sin(\theta_{ij})$, where we see that the theoretical maximum for this power flow is given by $P_{ij} = -V_i V_j b_{ij}$. Hence for a stealthy attack it is not enough to require that $a \in \text{Im}(H)$ but it is also essential to impose saturation limits on the attacked measurements, included in the set of constraints $\mathcal{C}$, in general reducing the set of valid attacks.

Varying operating conditions.

The power network is a dynamical system and its state is frequently changing. Thus it might be the case that the attacker has previously obtained a linear model $\tilde{H}$ for a state $\tilde{x}$ and the attack is performed only when the system is in a different state x where the linear approximation is H. Hence for small attack vectors or for cases where $x \approx \tilde{x}$, the residual will be small and the attack may pass undetected. For larger attacks, however, this might not hold.

These scenarios are the motivation for the study reported in the following section. In fact, the framework presented next can be used to characterize stealthy attacks computed with perturbed linear models.

3.5 Stealthy Deception Attacks with Perturbed Adversarial Knowledge

Here we consider the scenario where the attacker is performing an attack according to (3.13), but having only a partial or corrupted knowledge of the measurement model. Such knowledge may be obtained, for instance, by recording and analyzing data sent from the RTUs to the control center using suitable statistical methods. The corrupted measurement model may also correspond to an out-dated model or an estimated model using the power network topology, usual parameter values and uncertain operating point.

In the following analysis we provide bounds on the measurement residual under this kind of attack scenario. These bounds give some insights on what attacks may go undetected, given the model uncertainty. For the moment we assume there are no random errors in the measurements, i.e. $\epsilon = 0$.

Let the perturbed measurement model known by the attacker be denoted by $\tilde{H}$, such that

$$\tilde{H} = H + \Delta H. \tag{3.18}$$

For instance, one could have $H = \left.\dfrac{\partial h(x)}{\partial x}\right|_{x=x^*}$ while $\tilde{H} = \left.\dfrac{\partial h(x)}{\partial x}\right|_{x=\tilde{x}}$, with $x^* \neq \tilde{x}$, or even $\tilde{H} = H_{DC}$ as used in the experimental example at the end of this section. Additionally, consider the linear policy to compute attacks on the measurements to be $a = \tilde{H}c$, resulting in the corrupted set of measurements $z^a = z + a$. Recall the objectives of the attacker as defined in Section 3.4.1.

The objective of being undetected depends both on the desired bias on the flow measurements a and on the model uncertainty ΔH. The measurement residual under attack, $r^a = r(z^a)$, can be written as

$$r(z^a) = S(z + \tilde{H}c) = Sz + r_a, \tag{3.19}$$

which follows from (3.5) and $S = S^2$, since S is an orthogonal projector. Using (3.18) and the fact that $S = \mathbf{P}_{\mathrm{Ker}(H^\top)}$, we can rewrite the residual as

$$r(z^a) = S(z + Hc) + S\Delta Hc = S\Delta Hc. \tag{3.20}$$

We denote $r_a = S\Delta Hc$ as the residual due to the attack, since it only depends on c and ΔH. Furthermore, we see that $\|r_a\| \leq \|S\|\|\Delta H\|\|c\| = \|\Delta H\|\|c\|$, since S is an orthogonal projector, showing that the residual norm is linear in terms of the model uncertainty. However, this bound does not capture an important property of the sensitivity matrix S, *i.e.*, S is the orthogonal projector onto $\operatorname{Ker}(H^\top)$. To show this, assume $\tilde{H} = \Delta H$ for some nonzero δ, yielding $\Delta H = (1-\delta)H$. From the previous result we have $\|r_a\| \leq \|(1-\delta)H\|\|c\|$. However, since S is the orthogonal projector onto $\operatorname{Ker}(H^\top)$ and this subspace is the orthogonal complement of $\operatorname{Im}(H)$ we know that $r_a = S\Delta Hc = 0$. Therefore, although there is model uncertainty, the residual is still zero. This reasoning indicates that there is a geometrical meaning in the residual, since all the model perturbations ΔH spanning $\operatorname{Im}(H)$ will yield a zero residual. To further explore this property, we will make use of the so-called principal angles and projection theory described in (Galántai, 2006). The main results and definitions used in this work are now given.

Definition 3.5.1 ((Galántai, 2006)). *Let M_1 and M_2 be subspaces of $\mathbb{C}^m$. The smallest principal angle $\gamma_1 \in [0, \pi/2]$ between M_1 and M_2 is defined by*

$$\begin{aligned} \cos\gamma_1 = \max_{u\in M_1}\max_{v\in M_2} |u^H v| \\ \textit{subject to } \|u\| = \|v\| = 1 \end{aligned} \tag{3.21}$$

Lemma 3.5.2 ((Galántai, 2006)). *Let $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{m\times m}$ be orthogonal projectors of M_1 and M_2, respectively. Then the following holds*

$$\|\mathbf{P}_1\mathbf{P}_2\|_2 = \cos(\gamma_1) \tag{3.22}$$

Proposition 3.5.3. *Let γ_1 be the smallest principal angle between* $\operatorname{Ker}(H^\top)$ *and* $\operatorname{Im}(\tilde{H})$*. The residual increment due to a deception attack following the policy $a = \tilde{H}c$ satisfies*

$$\|r_a\|_2 \leq \cos\gamma_1 \|a\|_2. \tag{3.23}$$

Proof. Recall the so-called hat matrix defined by $K = HH^\dagger$, which is the orthogonal projector onto $\operatorname{Im}(H)$ and define $\tilde{K} = \mathbf{P}_{\operatorname{Im}(\tilde{H})} = \tilde{H}\tilde{H}^\dagger$. The residual under attack in (3.19) may be rewritten as

$$r_a = S\tilde{K}\tilde{H}c, \tag{3.24}$$

since $\tilde{K}\tilde{H} = \tilde{H}$. The residual norm can be upper bounded as

$$\|r_a\|_2 \leq \|S\tilde{K}\|_2\|\tilde{H}c\|_2 = \cos\gamma_1\|a\|_2, \tag{3.25}$$

where γ_1 is the smallest principal angle between $\operatorname{Ker}(H^\top)$ and $\operatorname{Im}(\tilde{H})$. □

Analyzing the example where $\tilde{H} = \Delta H$, we see that $\operatorname{Im}(\tilde{H}) = \operatorname{Im}(H)$ is orthogonal to $\operatorname{Ker}(H^\top)$. Hence the smallest principal angle between these subspaces is $\gamma_1 = \frac{\pi}{2}$, yielding $\|r_a\|_2 \leq \cos(\gamma_1)\|a\|_2 = 0$.

Thus we achieved a tighter bound that explores the geometrical properties of the residual subspace. In brief, γ_1 measures how close the subspaces $\operatorname{Ker}(H^\top)$ and $\operatorname{Im}(\tilde{H})$ are from each other. In order for the model uncertainty not to affect the residual, it is desired that $\operatorname{Ker}(H^\top)$ and $\operatorname{Im}(\tilde{H})$ are as close to orthogonal as possible. For some insights on the physical interpretation of this geometrical property, see the illustrative example in Section 3.5.2.

3.5.1 $\bar{\delta}-$Stealthy Attacks

Consider the measurement residual under attack in (3.19). Taking into account the random error vector ϵ we can rewrite the residual as

$$r(z^a) = S\epsilon + Sa. \tag{3.26}$$

The residual then has the following distribution $r(z^a) \sim \mathcal{N}(r_a, S)$. Note that due to the model uncertainties the residual may have a non-zero mean, which increases the chances of triggering an alarm in the BDD. Recall that one of the attacker's objective is to keep such probability as low as possible, i.e., $\|Wr(z^a)\|_p < \tau$. We now provide insights on how such objective may be fulfilled for the two BDD schemes presented in Section 3.2.3.

Detection probability.

Recall the hypothesis tests described in Section 3.2.3 where the nominal case with no bad data is denoted by H_0, while H_1 corresponds to the presence of bad data, including stealthy attacks. The detection probability is defined as the conditional probability of detecting bad data given that they are indeed present, i.e., $\mathbb{P}(H_1|H_1) = \mathbb{P}(\|Wr(z^a)\|_p > \tau | a \neq 0)$.

For the stealthy attacks $a \in \operatorname{Im}(H)$, the detection probability is given by $\mathbb{P}(H_1|H_1) = \mathbb{P}(\|WS\epsilon\|_p > \tau | a \neq 0) = \mathbb{P}(\|WS\epsilon\|_p > \tau | a = 0) = \alpha$, where α is the false-alarm rate. Thus the detection probability is independent of a and triggering an alarm for stealthy attacks has the same probability as giving a false-alarm.

On the other hand, for $a \notin \operatorname{Im}(H)$ the detection probability is a function of the attack vector a and the previous concept of stealthiness cannot be used, i.e. having $\mathbb{P}(H_1|H_1) = \alpha$. Therefore we consider instead the increase in the detection probability, $\delta(a)$, given by

$$\delta(a) = P(H_1|H_1) - \alpha, \tag{3.27}$$

and define the class of $\bar{\delta}-$stealthy attacks as $a \in \mathcal{U}_{\bar{\delta}}$ with $\mathcal{U}_{\bar{\delta}} = \{a \in \mathbb{R}^m : \delta(a) \leq \bar{\delta}\}$, where $\bar{\delta}$ represents the maximum increase in detection allowed by the attacker model.

Example 3.1
Consider a non-central chi-squares distribution, $\chi^2_k(\varphi)$, with $k = 5$ degrees of freedom. Under nominal conditions we have $\varphi = 0$ and this variable becomes chi-squares distributed. An hypothesis test was designed to detect changes on φ with false-alarm rate $\alpha = 0.05$, as described in Section 3.2.3 for the performance index test. Figure 3.4 illustrates the probability density function for $\varphi = 1$ (solid curve) and for $\varphi = 5$ (dashed curve). The detection probability in both cases is also depicted (shaded area), which is larger for $\varphi = 5$.

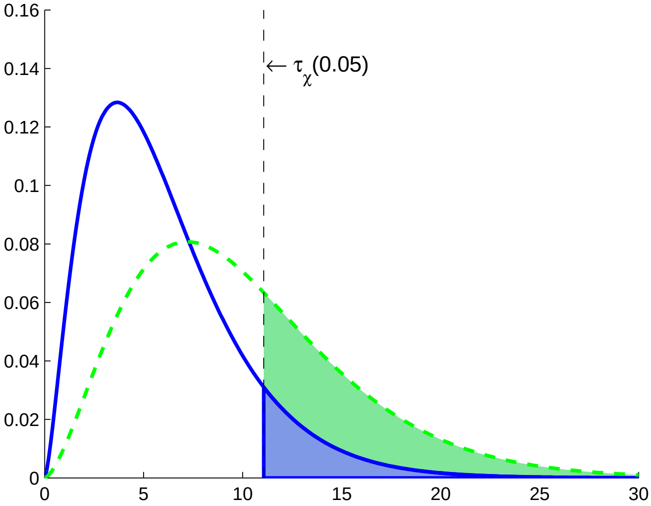


Figure 3.4: Probability distribution functions of non-central chi-squares distributions $\chi^2_k(\varphi)$ with $k = 5$ degrees of freedom and non-centrality parameter $\varphi = 1$ (solid) and $\varphi = 5$ (dashed). The shaded areas correspond to the detection probability for the corresponding φ, which is larger for $\varphi = 5$.

Performance index test

Recall that without any attack on the measurements we have $J(\hat{x}) \sim \chi^2_{m-n}$. Under attack the cost function $J_a(\hat{x}) = r(z^a)^\top r(z^a)$ will have the so-called *non-central chi-squares* distribution (Muirhead, 1982), due to the non-zero mean. We denote $J_a(\hat{x}) \sim \chi^2_{m-n}(\varphi)$ where $\varphi = \|Sa\|_2^2$. Recalling the relationship between the false-alarm probability α and the detection threshold $\tau_\chi(\alpha)$ in (3.7), in the presence of attacks we have

$$P(H_1|H_1) = \int_{\tau_\chi(\alpha)}^{\infty} g_\varphi(u)du = \alpha + \delta(\varphi), \tag{3.28}$$

with $g_\varphi(u)$ being the pdf of $\chi^2_{m-n}(\varphi)$.

For a given value of α, the class of $\bar{\delta}-$attacks is then characterized by the set of attacks for which $\varphi = \|Sa\|_2^2$ satisfies

$$\int_{\tau_\chi(\alpha)}^{\infty} g_\varphi(u)du \leq \alpha + \bar{\delta}. \tag{3.29}$$

A characterization of $\bar{\delta}-$attacks based on the geometric properties of H and $\tilde{H}$ is presented next based on the following assumption.

Assumption 3.5.1. *The change in detection probability, $\delta(\varphi)$, is increasing in φ.*

This seems to be a weak assumption given the example in Figure 3.4 and the fact that the mean and variance of $\chi^2_{m-n}(\varphi)$ increase as φ increases.

Proposition 3.5.4. *Given Assumption 3.5.1, α, and $\bar{\delta}$, an attack is $\bar{\delta}-$stealthy regarding the performance index test if the following holds*

$$\cos\gamma_1 \|a\|_2 \leq \sqrt{\bar{\varphi}(\alpha, \bar{\delta})} \tag{3.30}$$

where $\bar{\varphi}(\alpha, \bar{\delta})$ is the maximum value of φ for which (3.29) is satisfied and γ_1 is the smallest principal angle between $\mathrm{Ker}(H^\top)$ and $\mathrm{Im}(\tilde{H})$.

Proof. First note that from our assumption $\delta(\varphi)$ increases with φ. Therefore stealthy attack vectors satisfy $\|r_a\|_2 \leq \sqrt{\bar{\varphi}}$, as this implies by definition that $\varphi \leq \bar{\varphi}$ and $\delta(\varphi) \leq \bar{\delta}$. The rest of the proof follows from Prop. 3.5.3. □

Largest normalized residual test

Recall that the residuals without attack follow a normal distribution $r \sim \mathcal{N}(0, S)$, whereas under attack we have $r_a \sim \mathcal{N}(d, S)$ with $d = Sa$. Each element of the normalized residual vector then has distribution $r^N_{a_i} \sim \mathcal{N}(d^N_i, 1)$ with $d^N_i = D^{-1/2}_{ii} d_i$ being the bias introduced by the attack vector. Similarly as before, considering $\delta(d)$ and given α, the biases d^N_i for which the attacks are $del\bar{t}a-$stealthy satisfy the inequality

$$\int_{-\tau_{\mathcal{N}}(\alpha)}^{\tau_{\mathcal{N}}(\alpha)} g^{\mathcal{N}}_{d^N_i}(u)du \geq 1 - \alpha - \bar{\delta}, \tag{3.31}$$

with $g^{\mathcal{N}}_{d^N_i}(u)$ being the pdf of $r^N_{a_i}$.

Proposition 3.5.5. *Given α and $\bar{\delta}$ an attack is $\bar{\delta}-$stealthy regarding the largest normalized residual test if the following holds*

$$\|D^{-1/2}\|_2 \cos\gamma_1 \|a\|_2 \leq \bar{d}^N(\alpha, \bar{\delta}) \ , \tag{3.32}$$

where $\bar{d}^N(\alpha, \bar{\delta})$ is the maximum value of $\|d^N\|_\infty$ for which (3.31) is satisfied with $d^N_i = \|d^N\|_\infty$ and γ_1 is the smallest principal angle between $\mathrm{Ker}(H^\top)$ and $\mathrm{Im}(\tilde{H})$.

Proof. Clearly it is sufficient to require (3.31) to hold for $|d^N_i| = \|d^N\|_\infty$, as this corresponds to the worst-case bias. Note that the increase in alarm probability $\delta(d)$ increases with $|d^N_i|$ due to the symmetrical nature of $g^{\mathcal{N}}_{d^N_i}(u)$. Thus (3.31) reaches equality for $\|d^N\|_\infty = \bar{d}^N$ and a sufficient condition for (3.31) to hold is to have $\|d^N\|_\infty \leq \bar{d}^N$. Recalling $d^N = D^{-1/2}Sa$ and $\|\cdot\|_\infty \leq \|\cdot\|_2$, we conclude the attack is stealthy if $\|D^{-1/2}Sa\|_2 \leq \bar{d}^N$, which is satisfied by $\|D^{-1/2}\|_2\|Sa\|_2 \leq \bar{d}^N$. The rest follows from Proposition 3.5.3. □

The main result of this section is as follows:

Theorem 3.5.6. *Given the perturbed model $\tilde{H}$, the false-alarm probability α and the maximum admissible increase in alarm probability $\bar{\delta}$, an attack following the policy $a = \tilde{H}c$ is stealthy if*

$$\|a\|_2 \leq \beta(\alpha, \bar{\delta}) \ , \tag{3.33}$$

where $\beta(\alpha, \bar{\delta})$ is given by:

- $\beta(\alpha, \bar{\delta}) = \dfrac{\sqrt{\bar{\varphi}(\alpha, \bar{\delta})}}{\cos \gamma_1}$, *for the performance index test;*
- $\beta(\alpha, \bar{\delta}) = \dfrac{\bar{d}^N(\alpha, \bar{\delta})}{\|D^{-1/2}\|_2 \cos \gamma_1}$, *for the largest normalized residual test,*

where γ_1 is the smallest principal angle between $\mathrm{Ker}(H^\top)$ *and* $\mathrm{Im}(\tilde{H})$*.*

Proof. Assuming the BDD method is the performance index and taking $\beta(\alpha, \bar{\delta}) = \frac{\sqrt{\bar{\varphi}(\alpha, \bar{\delta})}}{\cos \gamma_1}$, the proof directly follows from Proposition 3.5.4. For the largest normalized residual, defining $\beta(\alpha, \bar{\delta}) = \frac{\bar{d}^N(\alpha, \bar{\delta})}{\|D^{-1/2}\|_2 \cos \gamma_1}$ the proof follows from Proposition 3.5.5. □

Note that in the former scenario, the designer of the BDD scheme chooses both the detection method as well as the false-alarm probability α. These elements are fixed and usually unknown to the attacker, who defines the maximum acceptable risk, $\bar{\delta}$ and has some knowledge of the power network $\tilde{H}$. The proposed framework could be used by system designers to analyze vulnerabilities in different attack scenarios, for instance scenarios in which the attacker knows the DC network model or only knows the topology of the power network. Both these scenarios are illustrated in the remainder of this section.

3.5.2 Illustrative Example

An interesting analysis is to understand what is the worst-case uncertainty for the attacker, ΔH, maximizing the orthogonality between $\mathrm{Im}(\tilde{H})$ and $\mathrm{Im}(H)$. This corresponds to maximizing the effect of the attack vector a on the measurement residual. From the attacker's view, this could lead to a set of robust attack policies. As for the control center this could be useful to implement security measures based on decoys, for instance. It is known that the network model used in the SE can be kept in the databases of the SCADA system with little protection. Thus a possible defensive strategy would be, for instance, to disseminate a perturbed model with fake but "genuine“ looking parameter values in the database which, if retrieved and used by an attacker, would produce large residuals and increase the detection of intelligent attacks.

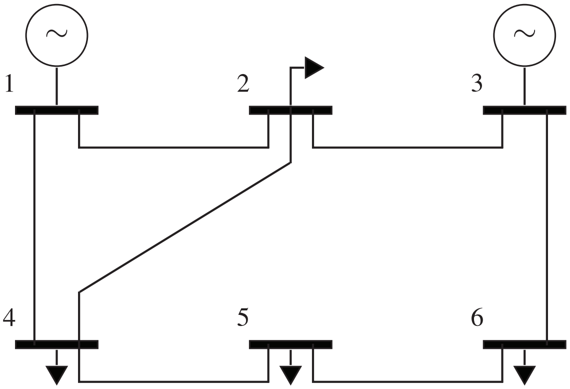


Figure 3.5: Power network with 6 buses

Table 3.1: Data of the network in Figure 3.5

Branch	From bus	To bus	Reactance (pu)	Parameter Error
$b1$	1	4	0.370	-20%
$b2$	1	2	0.518	+20%
$b3$	6	5	1.05	-20%
$b4$	6	3	0.640	-20%
$b5$	5	4	0.133	-20%
$b6$	4	2	0.407	-20%
$b7$	3	2	0.300	+20%

The first observation at this point is that it is of little interest to consider cases when only the maximum magnitude of the model perturbation is considered,*i.e.* $\|\Delta H\| \leq \omega$. Note that this formulation only tells us that the uncertainty is within a ball of radius ω from the nominal model H. Thus one can always choose a worst-case perturbation satisfying $\|\Delta H\| = \omega$ which is orthogonal to H, yielding $\|SK_\Delta\| = 1$. Hence scenarios where the uncertainty is more structured are of greater interest.

We now apply the previous results to the scenario where the attacker knows the exact topology of the network but has an error on the transmission line's parameters of ±20%. The detectability of attacks in this scenario is intimately related to the detectability of parameter or topology errors (Liu et al., 1992; Wu and Liu, 1989). Consider the power network in Figure 3.5 with the data in Table 3.1. The network shown in Figure 3.5 corresponds to the bus-branch model of a, possibly larger, power network computed by the EMS after analyzing which buses and branches are

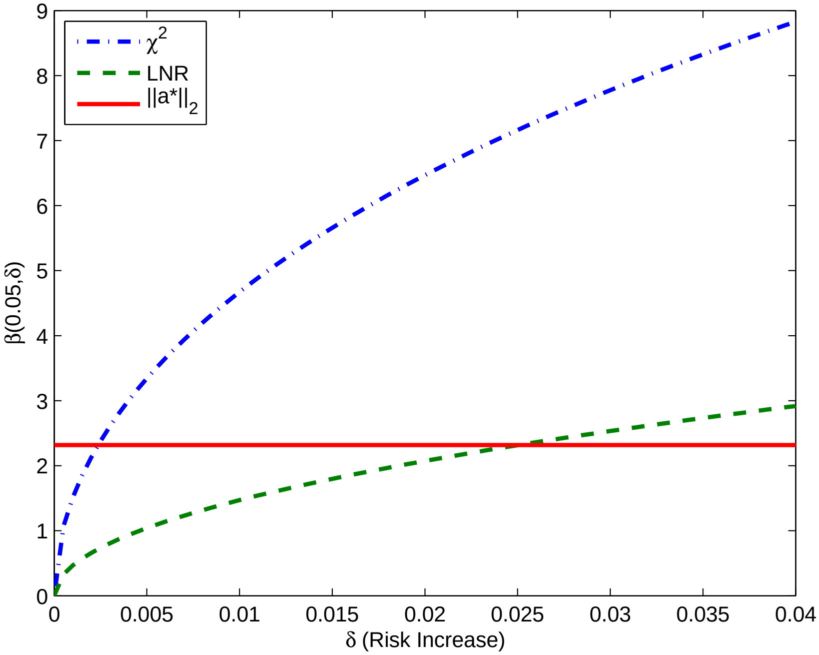


Figure 3.6: Attack stealthiness as a function of the detection risk. The solid line represents the 2-norm of the optimal attack vector a^* constrained by $a_{b_1} = 1$, where a_{b_1} is the power flow in branch b_1. The curves denoted as χ^2 and *LNR* represent the value of $\beta(0.05, \delta)$ for the performance index test and largest normalized residual test, respectively. From these results, we conclude that the *LNR* test is more sensitive to this kind of attacks.

energized, based on measurements from RTUs such as breaker status. This model is then used by the SE, together with the list of available measurements, to compute the measurement model. In this example we consider the linear case where $z = Hx$. The parameter errors in Table. 3.1 were computed so that $\cos(\gamma_1) = \|S\tilde{K}\|_2$ is maximized for errors up to $\pm 20\%$, corresponding to the worst-case uncertainty. This actually corresponds to the constrained maximization of a convex function, which was solved using the numerical solvers available in MATLAB.

In Figure 3.6 we show how the maximum 2-norm of a stealthy attack vector $\beta(\alpha, \delta)$ in terms of Theorem 3.5.6 varies with respect to the increase in the detection probability δ, for $\alpha = 0.05$. As it is seen, the performance index test allows for larger attacks than the largest normalized residual test. Since attacks following $a = \tilde{H}c$

have a similar meaning to multiple interacting bad data, this validates the known fact that largest normalized residual test is more robust to such bad data than the performance index test (Abur and Exposito, 2004). Note that the norm of the optimal attack vector in the sense of (3.15) when targeting the power flow between buses 1 and 4 is also shown. We see that such attack would have a small risk, even for the largest normalized residual.

3.5.3 Experiments on a SCADA EMS

The previous sections considered stealthy deception attacks on the SE based on linearized models and a new result was derived, characterizing a set of stealthy attack for given model perturbations. However, the results obtained so far do not clarify how sensitive SCADA EMS softwares are to these attacks or if a system operator should even care about these scenarios. Recall, for instance, that in the previous discussions the attack vector a was assumed to be sufficiently small, remaining unclear how large they can be while having the SE to converge and remaining stealthy.

In this section we present the results obtained by carrying out a stealthy deception attack on SCADA EMS software with the VIKING 40−bus benchmark described in Section 3.3. By analyzing these results, we hope to answer the previous open questions. Before analyzing the results, we briefly describe the experimental setup.

Experimental setup

The VIKING 40−bus benchmark with the default measurement configuration is considered. The measurements are available at each substation, which are kept in the software database. In these experiments, the data corruption was performed by directly changing the measurement data in the database. Thus the results obtained relate to deception attacks that were not detected or mitigated by standard IT security mechanisms.

Specific EMS components, such as SE and BDD, are configured with unitary weights for all the measurements. The SE solves the nonlinear weighted least-squares problem using the fast-decoupled algorithm with equality constraints (Wu, 1990), while the BDD algorithm uses the LNR test described in Section 3.2.3.

As in the numerical experiments in Section 3.4.3, the DC network model H_{DC} was used for the stealthy attack synthesis. Recall that this model is obtained by assuming all voltage magnitudes to be 1 pu and the phase-angles 0, while neglecting the line resistances and shunt admittances.

Attack scenario

To conduct our experiment we considered measurement number 33, corresponding to the active power flow on the tie-line between TROY and BLOO substations,

Table 3.2: Example: adding 100MW to target measurement 33

Measurement index, k	Normalized attack, $\bar{a}_k$	Correct value (MW), z_k^*	False value (MW), z_k^a
4	-1	1005.7041	905.7042
21	-0.7774	157.8541	80.1103
24	0.9665	507.7171	604.3638
27	2.7439	40.0006	314.3911
33	**1**	**-14.7971**	**85.2029**
62	0.7774	-123.3764	-45.6327
104	-0.9665	-334.8826	-431.5293

to be the target measurement that the attacker desires to corrupt. In order to do so without being detected, the attacker needs to perform a coordinated attack by corrupting the value of other power measurements. Following the framework presented in Section 3.4.2, the set of such malicious changes is encoded in the attack vector a, and z^a follows from (3.11).

Using H_{DC}, we computed the additive normalized attack vector required to stealthily change the target measurement by 1 MW, presented in Table 3.2. The several non-zero elements of the normalized attack have similar order of magnitude, which indicates that saturation limitations may not be too restrictive in this scenario, as the size of the bias in all the measurements will be roughly the same.

As seen in Table 3.2, this attack only corrupts 7 measurements in total, which are taken from 5 neighboring substations, namely TROY, BLOO, JUNE, MONR, and CROS. Hence we see that to stealthily attack a single measurement, a local coordinated attack suffices, even for such a large system. Additionally, as discussed in (Dán and Sandberg, 2010), note that usually all measurements within a given substation are gathered at a single RTU. This means that by breaking into the substation's RTU the attacker gains access to all those measurements, so we can argue that although 7 measurements need to be corrupted, only 5 RTUs need to be compromised.

In terms of the attack model in Section 3.4.1, this scenario can formulated with $\mathcal{G} = \{a \in \mathbb{R}^m : a_{33} = 1\}$, $\mathcal{U} = \mathrm{Im}(H_{DC})$ and $\mathcal{C} = \{a \in \mathbb{R}^m : a_i = 0, \forall i \in \mathcal{I}_p\}$, where $\mathcal{I}_p$ is the set of protected measurements, including pseudo-measurements.

Experimental results

The normalized attack vector $\bar{a}$, whose non-zero entries are shown in Table 3.2, was used to corrupt the measurement data according to the attacker's objective. For instance, in Table 3.2 we can see the correct value of the compromised measurements, denoted by z^*, and the false values sent to the control center, z^a, when the objective was to induce a bias of 100MW in the target measurement, having

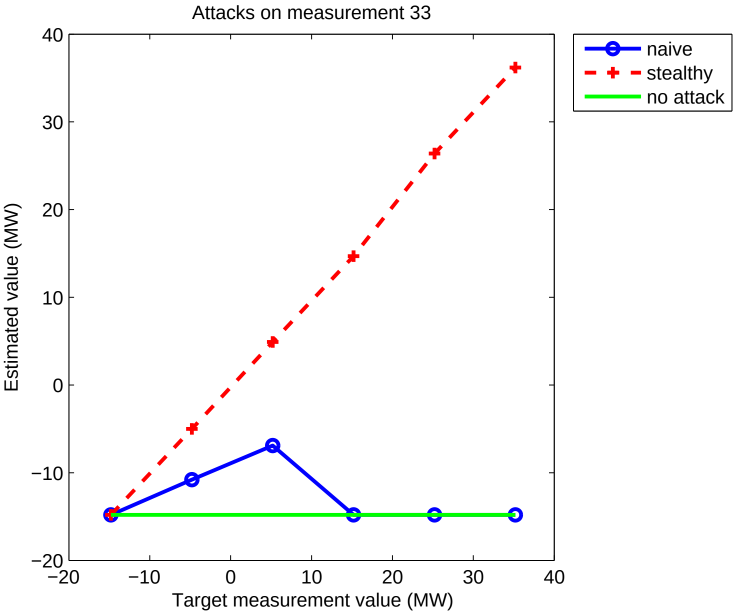


Figure 3.7: Stealthy deception attack

$z^a = z^* + 100\bar{a}$.

In Figure 3.7 we show the results obtained by performing stealthy deception attacks as described before and naive deception attacks where only the target measurement is compromised. In both cases, the bias in the target measurement was sequentially increased by 10MW at each step. From these results we see that the naive attack was undetected up to a bias of 20MW, while for bias above 30MW this attack was detected and the compromised measurement removed. The coordinated stealthy attack, however, remained undetected for all the bias values showed in the figure. Furthermore we see that the naive attack did not influence the estimate as much as the stealthy one. For the stealthy attacks relationship between the false and the estimated values is an almost unitary slope.

Table 3.3 shows the results obtained for large bias, where the attacks were performed sequentially with steps of 50MW. We observe that the stealthy attacks were successful with no BDD alarm triggered up to a bias of 150MW, beyond which the SE no longer converged which violates condition 1) of the attacker model in Section 3.4.1.

Although the SE did not converge for attacks above 200MW, it is still surprising to see that attacks based on the linearized model as large as 150MW are successful. To better understand what such quantity indicates, note that the nominal value of

Table 3.3: Results from the stealthy attack for large bias

Target bias, a_{33}	False value (MW), z^a_{33}	Estimate (MW), $\hat{z}^a_{33}$	#BDD Alarms	#CA Alarms
0	-14.8	-14.8	0	2
50	35.2	36.2	0	2
100	85.2	86.7	0	10
150	135.2	137.5	0	27
200	185.2	-	-	-

the targeted tie-line is 260MW. Thus the attack was able to induce a bias of more than 50% of the nominal value, which reveals that the SCADA EMS software is indeed sensitive to stealthy deception attacks. Furthermore, notice that the number of warnings given by the CA component increase with the size of the attack. Similar behavior is expected in other EMS software, given that the theory behind the SE and BDD components is quite standard in the literature, as described in Section 3.2.

Note that these results were achieved with a simplified linear model where several parameters were disregarded, including the correct operating conditions and cross-coupling effects between active and reactive measurements. However, the assumptions in this scenario are still rather strong. Recall that the attacker was assumed to have a large amount of resources such as a rather detailed knowledge regarding the DC network model, the available measurements, and the pseudo-measurements, and access to several RTUs. Most likely, an attacker with such resources could find easier alternative attacks on the power network than the one considered in this chapter.

Nevertheless, these vulnerabilities do exist and can be exploited by attackers with the required resources. Moreover, as described in Section 2.4, the vulnerabilities on the SE could propagate to other components such as the CA and OPF and hence affect the operators' actions. It is yet unclear how these would affect the power network operation.

The increased number of CA warnings could lead the operator to take corrective actions, as the CA warnings indicate that the system does not meet the reliability criteria. On the other hand, the OPF would give the operator misleading recommendations, computed based on the compromised state estimate. The latter scenario is further discussed in Chapter 4.

3.6 Summary

This chapter presented a comprehensive framework to analyze and study stealthy deception attacks specifically targeting the SE and BDD components of SCADA EMS software through measurement data corruption. The proposed framework is able to include different attacker and attack cost models and characterizes existing limitations on stealthy attacks depending on the available model knowledge. In particular, we proposed a novel security metric for each measurement, corresponding to the minimum number of sensors needed to have stealthy attacks targeting each measurement. The different concepts and results of the framework are illustrated through numerical examples using the VIKING 40−bus benchmark.

To validate this framework, we took a novel approach and conducted a set of deception attacks to a SCADA EMS software. In spite of the power network being nonlinear, the results obtained by this experiment show that computations based on linear models provide valid stealthy attacks. These attacks successfully corrupt the target measurements without triggering any BDD alarms.

The results also indicate that linear models can be used for large attacks as well, although otherwise could expected. However, in addition to the measurement model, information concerning pseudo-measurements and saturation limits is needed for a successful stealthy attack. This fact can be used by the utility to devise new protection schemes. Additionally, the proposed security metric may help security experts to prioritize and rank system components when deploying protective schemes.

Chapter 4

Cyber-Security of Optimal Power Flow in Power Systems

The previous chapter analyzed the SE and BDD components and characterized existing vulnerabilities that can be exploited to perform stealthy deception attacks on the measurement data. These vulnerabilities propagate to other EMS components such as the OPF and CA, since these are based on the output from the SE, and may affect the power network operation. How the stealthy attacks on measurements affect the OPF, CA, operator, and power network operation are still relatively open questions.

In this chapter we consider the behavior of the OPF component given the corrupted measurements and discuss scenarios that might lead the operator to use the misleading OPF recommendations, thus affecting the power network.

The outline of this chapter is as follows. An overview of the related work and our contributions are given in Section 4.1. Section 4.2 describes a simplified OPF algorithm based on the DC model, namely the DC-OPF. The characterization of the impact of stealthy attacks on the power network and a novel impact-aware security metric are given in Section 4.3, followed by a numerical example. A summary of the chapter is presented in Section 4.4.

4.1 Contributions and Related Work

As illustrated in Figure 3.1, stealthy deception attacks bypassing the BDD may affect other components of the EMS. In particular the stealthy data attacks may disturb the OPF algorithm and, consequently, the human operator's decision regarding supervisory changes to the power network operation.

The economic impact of data corruption attacks has been investigated recently for electricity market applications (Jia et al., 2011; Xie et al., 2010). These approaches considered the linearized version of the DC-OPF or Economic Dispatch and provided several attack heuristics to tamper with the electricity markets while remaining undetected by conventional BDD schemes. Recently the impact of a more restricted class of data attacks corrupting only load and flow measurements was also analyzed for a linear Security-Constrained Economic Dispatch problem under a game-theoretic perspective in (Yuan et al., 2011). In these approaches the presence of a human operator was neglected, hence the compromised control actions were always applied to the system.

Our work analyzes the behavior of the DC-OPF, formulated as a Quadratic Programming problem, under corrupted estimates resulting from stealthy attacks. No market application is considered; instead we focus on how the corrupted estimates may affect the operator's decisions and the possible direct economic consequences in terms of increase generation costs and resistive losses. We further consider that the operator makes a binary decision of either closing the loop over the DC-OPF recommendation or not taking any control action based on the expected economic profit.

Our first contribution builds on the KKT conditions and provides an analytical characterization of the perturbed DC-OPF solution assuming that the corrupted estimates satisfy certain constraints. Using these expressions, we discuss under what conditions a human operator has the incentive to consider and apply the compromised DC-OPF recommendation. The discussion follows by analyzing the economic impact on the network operation if the operator decides to close the control loop over the corrupted measurements. An analytical example is discussed, illustrating the concepts of the first contribution and motivating our second one.

In our second contribution, an impact aware security metric is proposed using the novel analytical expressions derived. A possible improvement to the security index proposed by Sandberg et al. (2010), so that the attack impact is considered, is also discussed. The usefulness of our novel impact aware security metric is illustrated through a numerical example, indicating that these contributions may be helpful to system security designers, namely for secure sensor allocation, as discussed in (Bobba et al., 2010) and (Dán and Sandberg, 2010).

4.2 Preliminaries

4.2.1 Simplified Optimal Power Flow

In this work a simplified OPF problem is considered, namely the DC-OPF using the DC network model explained in Section 3.2.1. Let N, N_g, and N_b be the number of buses, generator buses, and transmission lines in the power network, respectively. The variables considered in the DC-OPF problem are:

- $P^d \in \mathbb{R}^N$: the active power demand;
- $P^g \in \mathbb{R}^{N_g}$: the active power generation;
- $\theta \in \mathbb{R}^{N-1}$: the phase-angle at each bus, except the reference bus, for which $\theta_1 = 0$;

- $P^f \in \mathbb{R}^{N_b}$: the active power flow on each transmission line.

The active power demand P^d is supplied to the DC-OPF as a known parameter, while the power generated at all generator buses, P^g, are the decision variables constrained by $\overline{P^g} \geq P^g \geq 0$. In the DC network model the power equations (3.2) provide a linear relation between the phase-angles and the power demand, generation, and flows:

$$\begin{bmatrix} C_g P^g - P^d \\ P^f \end{bmatrix} = H_{DC}\theta = \begin{bmatrix} H_i \\ H_f \end{bmatrix} \theta, \tag{4.1}$$

where $H_{DC} \in \mathbb{R}^{N+N_b \times N-1}$ as in Section 3.2.1, $C_g \in \mathbb{R}^{N \times N_g}$ is the bus to generator incidence matrix, mapping the generators to the respective buses, and $C_g P^g - P^d$ are the power injections.

Assuming the power network is connected, an invertible matrix $\bar{H}_i$ is obtained by removing from H_i the row corresponding to the chosen slack bus, and so θ is obtained as a function of P^g and P^d. The power generated by the slack bus is determined so that the demand is met, meaning it is determined by solving

$$\mathbf{1}^\top P^g + \mathbf{1}^\top P^d = 0 \tag{4.2}$$

in the lossless case. The power flows can then be written as $P^f = G_g P^g + G_d P^d$, where G_g and G_d are obtained from $\bar{H}_i$, (4.1), and (4.2).

Thermal limitations on the transmission lines introduce operation limits on the power flows, $|P^f| \leq \overline{P^f}$. At the same time, there is an operation cost associated to each generator k, $c_k(P_k^g) = c_{k2}(P_k^g)^2 + c_{k1}P_k^g + c_{k0}$ with $c_{k2} > 0$. The purpose of DC-OPF is then to minimize the total generation cost subject to the operation limits of the transmission lines and generators, which can be formulated as the following optimization problem

$$\min_{P^g} c(P^g) := \sum_{k=1}^{N_g} c_k(P_k^g) \tag{4.3}$$

subject to

$$h(P^g, P^d) = \mathbf{1}^\top P^g + \mathbf{1}^\top P^d = 0 \tag{4.4}$$

$$f(P^g, P^d) = F_g P^g + F_d P^d + F_0 \leq 0, \tag{4.5}$$

with

$$F_g = \begin{bmatrix} G_g \\ -G_g \\ I \\ -I \end{bmatrix}, \quad F_d = \begin{bmatrix} G_d \\ -G_d \\ 0 \\ 0 \end{bmatrix}, \quad F_0 = \begin{bmatrix} -\overline{P^f} \\ -\overline{P^f} \\ -\overline{P^g} \\ 0 \end{bmatrix}.$$

The Lagrangian function for this problem becomes

$$L(P^g, \nu, \lambda) = c(P^g) + \nu(\mathbf{1}^\top P^g + \mathbf{1}^\top P^d) + \lambda^\top (F_g P^g + F_d P^d + F_0),$$

where ν and $\lambda \geq 0$ are the dual variables. In the following, we assume the DC-OPF problem is always feasible.

Optimal solution

For given P^d let us denote the optimal solution of the DC-OPF by

$$P^{g*} = \Omega(P^d) \tag{4.6}$$

and the associated nominal optimal cost by $c^* = c(P^{g*})$. The optimal power flows become $P^{f*} = G_g P^{g*} + G_d P^d$. Recalling that $c_{k2} > 0$, the DC-OPF corresponds to the minimization of a positive definite quadratic function subject to linear constraints and is therefore a strictly convex problem, for which a feasible optimal solution satisfies the the KKT conditions (Boyd and Vandenberghe, 2004):

$$\begin{aligned}
&0 = \nabla c(P^{g*}) + \nabla h(P^{g*}, P^d)^\top \nu^* + \nabla f(P^{g*}, P^d)^\top \lambda^* \\
&0 = h(P^{g*}, P^d) \\
&0 = \lambda_i^* f_i(P^{g*}, P^d), \quad \forall i = 1, \ldots, N_f \\
&0 \leq \lambda^*,
\end{aligned}$$

where $N_f = 2(N_b + N_g)$ is the number of inequality contraints.

According to the KKT conditions, only the dual variables associated with active constraints are nonzero in the optimal solution. We denote the number of active and inactive inequality constraints as N_1 and $N_0 = N_f - N_1$, respectively.

Rewriting the quadratic objective function as

$$c(P^g) = \frac{1}{2} {P^g}^\top Q P^g + R^\top P^g + C_0 \tag{4.7}$$

and denoting $H_1 \in \mathbb{R}^{N_1 \times N_f}$ and $H_0 \in \mathbb{R}^{N_0 \times N_f}$ as the "selector matrices" selecting the active and inactive constraints at the optimal solution, respectively, the KKT conditions become

$$\begin{bmatrix} Q & F_g^\top & \mathbf{1} \\ \mathbf{1}^\top & 0 & 0 \\ H_1 F_g & 0 & 0 \\ 0 & H_0 & 0 \end{bmatrix} \begin{bmatrix} P^{g*} \\ \lambda^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -R \\ -\mathbf{1}^\top (P^d) \\ H_1(-F_d P^d - F_0) \\ 0 \end{bmatrix},$$

which we rewrite as

$$V \begin{bmatrix} P^{g*} \\ \lambda^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -R \\ -\mathbf{1}^\top P^d \\ H_1(-F_d P^d - F_0) \\ 0 \end{bmatrix}. \tag{4.8}$$

Proposition 4.2.1. *For any feasible optimal solution $P^{g*} = \Omega(P^d)$, λ^*, and ν^*, the corresponding matrix $V \in \mathbb{R}^{N_g+N_f+1 \times N_g+N_f+1}$ is invertible.*

Proof. Since the DC-OPF is a strictly convex optimization problem, there exists a unique optimal solution. Furthermore, since all inequality constraints are linear, strong duality holds and the KKT conditions (4.8) are necessary and sufficient for any feasible solution to be optimal. Hence (4.8) has a single solution, which requires V to be invertible. □

Nominal optimal operation

As discussed in Section 2.4, in general the full state of the power network is not directly available to the operator. Instead, the network state is estimated based on a large amount of measurements and a known measurement model using the algorithms described in Section 3.2.

Denoting $\hat{P}^g$, $\hat{P}^d$, and $\hat{P}^f$ as the estimated power generation, demand, and flows, the current *estimated operation cost* is

$$\hat{c} = c(\hat{P}^g) \tag{4.9}$$

and the OPF problem solved is in fact

$$\begin{aligned} \min_{P^g} \; & c(P^g) \\ \text{s.t.} \; & \mathbf{1}^\top P^g + \mathbf{1}^\top \hat{P}^d = 0 \\ & F_g P^g + F_d \hat{P}^d + F_0 \leq 0. \end{aligned}$$

Let us consider a system in which the estimated demand equals the true demand $\hat{P}^d = P^d$ (i.e., measurements are accurate). Applying $P^{g*} = \Omega(\hat{P}^d)$ to the system would then result in $\hat{P}^g = P^{g*}$ and $\hat{P}^f = P^{f*}$. We refer to this system as the system in *nominal optimal operation.*

Assumption 4.2.1. *The system operates in optimality, that is, $\hat{P}^g = P^{g*}$ and $\hat{P}^f = P^{f*}$ for the given $\hat{P}^d = P^d$.*

4.3 Cost Impact of Stealthy Data Attacks

Recall that the input parameters to the DC-OPF are obtained either from direct measurement or through state estimation and are therefore vulnerable to stealthy deception attacks the measurement data.

Now consider the case where the data attack illustrated in Figure 3.1 has been performed such that the corrupted estimates become

$$\hat{P}^g_a = P^{g*} + a_g = \hat{P}^g + a_g \tag{4.10}$$

$$\hat{P}^f_a = P^{f*} + a_f = \hat{P}^f + a_f \tag{4.11}$$

$$\hat{P}^d_a = P^d + a_d = \hat{P}^d + a_d \tag{4.12}$$

where a_g, a_d, and a_f are the corrupted data added to the measurements so that they fulfill (4.1) for some θ. That is, the attack is undetectable using standard bad-data detection schemes based on the DC network model, see Chapter 3. Recall that although these attacks are feasible, the consequences on the OPF and operator still remain unknown. In this section we address this issue.

Remark 4.3.1. *Given the corrupted estimate $\hat{P}_a^g$, the operator believes that the power network is operating at the* estimated operation cost *$c(\hat{P}_a^g)$.*

After receiving the corrupted measurements and computing the state estimates $\hat{P}_a^d$, the operator solves the DC-OPF problem and obtains the corresponding optimal solution $\hat{P}_a^{g*} = \Omega(\hat{P}_a^d)$. Before characterizing the DC-OPF solution given the corrupted measurements, we make the following assumption.

Assumption 4.3.2. *The data corruptions a_g, a_d, and a_f are sufficiently small so that the active constraints for $P^{g*} = \Omega(P^d)$ remain the same for $\hat{P}_a^{g*} = \Omega(\hat{P}_a^d)$.*

Conditions enforcing the above assumption to hold may be found in Section 4.3.4.

In the remainder of this section we discuss the consequences of the data corruption attack. First we characterize the data attack impact on the DC-OPF solution and under what conditions the operator may decide to apply the generation profile recommended by the DC-OPF under data attack. Assuming the operator accepts the DC-OPF recommendation, the discussion then proceeds by examining the true economical losses of that decision.

4.3.1 Consequences on the DC-OPF Solution

The DC-OPF solution given $\hat{P}_a^d$ can be computed using the KKT conditions in (4.8). Furthermore, based on Assumption 4.3.2, the difference in the optimal solutions $\hat{P}_a^{g*} = \Omega(\hat{P}_a^d)$ and $P^{g*} = \Omega(P^d)$ is given by

$$\begin{bmatrix} \hat{P}_a^{g*} - P^{g*} \\ \hat{\lambda}_{+_a}^* - \lambda_+^* \\ \hat{\lambda}_{-_a}^* - \lambda_-^* \\ \hat{\nu}_a^* - \nu^* \end{bmatrix} = V^{-1} \begin{bmatrix} 0 \\ \mathbf{1}^\top \\ H_1^+ G_d \\ -H_1^- G_d \\ 0 \\ 0 \end{bmatrix} a_d = \begin{bmatrix} T_g \\ T_+ \\ T_- \\ T_\nu \end{bmatrix} a_d, \tag{4.13}$$

where the invertibility of V follows from Proposition 4.2.1. Thus we can write

$$\hat{P}_a^{g*} - P^{g*} = T_g a_d, \tag{4.14}$$

where $T_g \in \mathbb{R}^{N_g \times N}$ is a linear mapping from changes in the loads to changes in the optimal generation profile.

At this point, the operator believes the power network can be operated at the *estimated optimal operation cost $c(\hat{P}_a^{g*})$* if the DC-OPF recommendation is applied.

Estimated re-dispatching profit

Given the corrupted power generation estimates $\hat{P}_a^g$, current *estimated operation cost* computed by the operator is $c(\hat{P}_a^g)$. Running the DC-OPF based on the corrupted load estimates $\hat{P}_a^d$ will provide the operator with the estimated optimal operation cost $c(\hat{P}_a^{g*})$. The difference between the estimated operation cost $c(\hat{P}_a^g)$ and the estimated optimal operation cost $c(\hat{P}_a^{g*})$ corresponds to the *estimated re-dispatching profit* if the power generation is re-dispatched according to the DC-OPF, which we now define using (4.10) and (4.14).

Definition 4.3.1 (Estimated Re-Dispatching Profit)**.** *The* estimated re-dispatching profit *is defined as*

$$\hat{\mathcal{P}}_a \triangleq c(\hat{P}_a^g) - c(\hat{P}_a^{g*}). \tag{4.15}$$

Observation 4.3.3. *Under Assumption 4.2.1 and Assumption 4.3.2, the* estimated re-dispatching profit *is given by*

$$\hat{\mathcal{P}}_a = c(\hat{P}^{g*} + a_g) - c(P^{g*} + T_g a_d) \approx \nabla c^{*\top}(a_g - T_g a_d). \tag{4.16}$$

Proof. The proof comes directly from the assumptions, (4.10), and (4.14). □

Note that $\hat{\mathcal{P}}_a$ is a quadratic function of the current optimal solution, P^{g*}, and the data corruptions added to the measurements of generation and demand, a_g and a_d, for which we provide a linear approximation.

Since the DC-OPF active and inactive constraints at optimality remain the same after the small data corruptions, given Assumption 4.3.2, we then conclude $\hat{\mathcal{P}}_a \geq 0$. A large value of $\hat{\mathcal{P}}_a$ can make the operator update the generator set-points, as there seems to be an incentive to do so. Note however that both these cost might be fictitious, since the estimates have been corrupted.

4.3.2 Consequences on the Physical Network

Consider that the operator decides to apply the generation profile $\hat{P}_a^{g*} = \Omega(\hat{P}_a^d)$ recommended by the DC-OPF under attack. In reality, the power demand may be different from the respective estimate, i.e. $P^d \neq \hat{P}^d$. This occurs for any data corruption attack with $a_d \neq 0$. Therefore there might exist a mismatch between the demand and generation, which has to be compensated by the slack generator so that the power balance equation (4.2) is satisfied. Without loss of generality, choosing generator 1 as the slack, the power generated by this bus is then a function of the power imbalance

$$P_{a,1}^{g*} = -\sum_{i=2}^{N_g} \hat{P}_{a,i}^{g*} - 1^\top P^d. \tag{4.17}$$

Hence the real generation profile after attack is $P_a^{g*} = [P_{a,1}^{g*}\ \hat{P}_{a,2}^{g*}\ \dots\ \hat{P}_{a,N_g}^{g*}]^\top$, yielding a *true operation cost* $c(P_a^{g*})$.

True re-dispatching profit

Assuming the operator applies the DC-OPF recommendation, the *true re-dispatching profit* due to the data corruption attack is defined as follows.

Definition 4.3.2 (True Re-Dispatching Profit)**.** *The* true re-dispatching profit *is defined as*

$$\mathcal{P}_a \triangleq c(P^{g*}) - c(P_a^{g*}). \tag{4.18}$$

A negative $\mathcal{P}_a$ value implies that the real operation cost is higher due to the attack. A positive value, however, would mean that the network is operating outside the feasible set considered by the DC-OPF. The proof to this statement comes from the convexity of the DC-OPF and the optimality principle. Such event might be possible since the constraints in the DC-OPF are soft constraints that guarantee a given safety margin. Thus $\mathcal{P}_a$ being positive implies a reduced safety margin in the current state.

Recalling (4.2), (4.14), and (4.17) , we can compute the true power generation difference $P_a^{g*} - P^{g*}$ as

$$P_a^{g*} - P^{g*} = \begin{bmatrix} 0 & -\mathbf{1}^\top \\ 0_{N_g-1\times 1} & I_{N_g-1} \end{bmatrix} T_g a_d,$$

which we rewrite as

$$P_a^{g*} - P^{g*} = MT_g a_d. \tag{4.19}$$

Observation 4.3.4. *Given Assumption 4.2.1 and Assumption 4.3.2, the* true re-dispatch profit *can be computed as*

$$\mathcal{P}_a = c(P^{g*}) - c(P^{g*} + MT_g a_d) \approx -\nabla c^{*\top} MT_g a_d. \tag{4.20}$$

Proof. Recalling Definition 4.3.2 and using (4.19) we have $\mathcal{P}_a \approx -\nabla c^{*\top}(P_a^{g*} - P^{g*}) = -\nabla c^{*\top} MT_g a_d$. □

In the next section we consider a simple analytical example to illustrate the discussion in the current section.

4.3.3 Analytical Example

Formulating the DC-OPF problem for the power network in Figure 4.1, we obtain the following optimization problem:

$$\min_{P^g} c_{12}(P_1^g)^2 + c_{11}P_k^g + c_{10} + c_{22}(P_2^g)^2 + c_{21}P_2^g + c_{20}$$

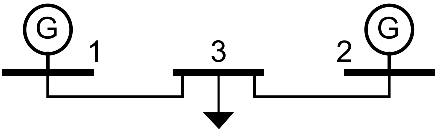


Figure 4.1: Three bus network with two transmission lines. The generations are $P_1^g, P_2^g \geq 0$ and the load is $P_3^d \leq 0$.

subject to

$$\begin{aligned} P_1^g + P_2^g + P_3^d &= 0 \\ P_{13}^f = P_1^g &\leq \overline{P^f}_{13} \\ P_{23}^f = P_2^g &\leq \overline{P^f}_{23} \\ -P_{13}^f = -P_1^g &\leq \overline{P^f}_{13} \\ -P_{23}^f = -P_2^g &\leq \overline{P^f}_{23}. \end{aligned}$$

We now consider the DC-OPF solution in a given scenario where we assume there are no saturated tie lines, namely $|P^{f*}| < \overline{P^f}$.

Nominal solution

Applying the KKT conditions (4.8) to this case, we obtain the following optimal generation profile

$$\begin{aligned} P_1^{g*} &= \frac{-c_{11} + c_{21} - 2c_{22}P_3^d}{2(c_{12} + c_{22})} \\ P_2^{g*} &= \frac{c_{11} - c_{21} - 2c_{12}P_3^d}{2(c_{12} + c_{22})} \end{aligned}$$

where we see the generated power depends on the generation costs. We now analyze the consequences of data corruption for this nominal operation scenario under Assumption 4.3.2.

Data corruption attack

Let us consider that the measurements are corrupted as in (4.10)–(4.12). The optimal generation profile, given the corrupted load estimates $\hat{P}_a^d$, is obtained by solving (4.8). For this scenario the solution is

$$\hat{P}^{g*}_{a,1} = \frac{-c_{11} + c_{21} - 2c_{22}\hat{P}^{d}_{a,3}}{2(c_{12} + c_{22})}$$
$$\hat{P}^{g*}_{a,2} = \frac{c_{11} - c_{21} - 2c_{12}\hat{P}^{d}_{a,3}}{2(c_{12} + c_{22})}$$

and the difference to the previous optimal generation profile P^{g*} is

$$\hat{P}^{g*}_{a} - P^{g*} = Ta_d = \frac{-1}{(c_{12} + c_{22})}\begin{bmatrix} c_{22} \\ c_{12} \end{bmatrix} a_d. \tag{4.21}$$

The *estimated re-dispatching profit* in (4.16) can be computed as

$$\hat{\mathcal{P}}_a \approx \nabla c^{*\top}(a_g - T_g a_d). \tag{4.22}$$

Similarly, the *true re-dispatching profit* due to the data corruption attack given by (4.20) can be computed as

$$\mathcal{P}_a \approx -\nabla c^{*\top} M T_g a_d. \tag{4.23}$$

A similar analysis is possible for other operating conditions, i.e. active set of constraints, possibly resulting in different matrices T_g and M.

Illustrative scenario

To illustrate the previous discussion, we now present two particular data attack scenarios based on the example network in Figure 4.1.

Consider P_1 acts as the slack bus and recall Assumption 4.2.1, which states that the system operates under optimality before the data attack. Furthermore, assume no lines are saturated and consider the data attack $a = \begin{bmatrix} a_g^\top & a_f^\top & a_d^\top \end{bmatrix}^\top$.

Scenario 1: ($c_{22} \gg c_{12}$) In this scenario the marginal cost of generator 2 is considered to be much higher than that of generator 1, hence in optimality it would be expected that an increase in the load demand would be compensated mainly by generator 1. Indeed using (4.21) we have

$$\hat{P}^{g*}_{a,1} - P^{g*}_{1} = \frac{-c_{22}}{(c_{12} + c_{22})} a_d \approx -a_d$$
$$\hat{P}^{g*}_{a,2} - P^{g*}_{2} = \frac{-c_{12}}{(c_{12} + c_{22})} a_d \approx 0,$$

meaning that the DC-OPF compensates small load changes solely through the cheapest bus, which happens to be the slack bus. Recalling that in open-loop, i.e.

without the DC-OPF, load changes are compensated by the slack bus, a direct consequence is that the true and estimated generator profile after applying the DC-OPF's recommendation are the same, as we can see from (4.21) and (4.19)

$$P_a^{g*} - P^{g*} = MT_g a_d = \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} a_d = 0.$$

Hence for all attacks $\mathcal{P}_a = 0$, i.e. there is no economic impact even if the DC-OPF solution is applied.

Scenario 2: $(c_{22} \ll c_{12})$ As opposed to the previous scenario, here the marginal cost of generator 2 is the lowest and hence we have

$$\begin{aligned} \hat{P}_{a,1}^{g*} - P_1^{g*} &= \frac{-c_{22}}{(c_{12}+c_{22})} a_d \approx 0 \\ \hat{P}_{a,2}^{g*} - P_2^{g*} &= \frac{-c_{12}}{(c_{12}+c_{22})} a_d \approx -a_d, \end{aligned}$$

indicating that small load changes are compensated by DC-OPF through the cheapest generator, which in this case is not the slack bus. Thus the DC-OPF and open-loop load compensations differ, possibly resulting in economic incentives to use the DC-OPF recommendation, as seen in the estimated re-dispatching profit

$$\hat{\mathcal{P}}_a \approx \nabla c^{*\top}(a_g - T_g a_d) = \nabla c^{*\top} \begin{bmatrix} a_{g,1} \\ a_{g,2} + a_d \end{bmatrix}.$$

The economic incentives to use the DC-OPF recommendation, the estimated re-dispatching profit $\hat{\mathcal{P}}_a$, are now analyzed for two different attacks. Suppose that all the measurements are available. It can be shown that in this case there are two 3−sparse attack patterns, namely attacks on measurements $\{a_{g,1},\, a_{f,13}\, a_d\}$ and $\{a_{g,2},\, a_{f,23}\, a_d\}$. Furthermore, given the DC power flow equations, $P_{13}^f = P_1^g$, $P_{23}^f = P_2^g$, and $P_3^d + P_{13}^f + P_{23}^f = 0$, these attacks are constrained by $a_{g,1} = a_{f,13} = -a_d$ and $a_{g,2} = a_{f,23} = -a_d$, respectively.

Consider the following normalized 3−sparse attack on $\{a_{g,1},\, a_{f,13}\, a_d\}$,

$$a^1 = \begin{bmatrix} a_{g,1} \\ a_{f,13} \\ a_d \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

Thus for a^1 we have the following estimated profit

$$\hat{\mathcal{P}}_a(a^1) \approx \nabla c^{*\top} \begin{bmatrix} a_{g,1} \\ a_{g,2} + a_d \end{bmatrix} = \nabla c^{*\top} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Considering now the other normalized 3−sparse attack on $\{a_{g,2},\, a_{f,23}\, a_d\}$,

$$a^2 = \begin{bmatrix} a_{g,2} \\ a_{f,23} \\ a_d \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

The estimated profit is then given by

$$\hat{\mathcal{P}}_a(a^2) \approx \nabla c^{*\top} \begin{bmatrix} a_{g,1} \\ a_{g,2} + a_d \end{bmatrix} = \nabla c^{*\top} \begin{bmatrix} 0 \\ 1 + 0 - 1 \end{bmatrix} = 0.$$

Thus we conclude that only the deception attack a^1 may lead the operator to re-dispatch the power generation, since the attack indicated that generator 1, the most expensive, compensated the fictitious load increase. On the other hand the attack a^2 will have no impact in the power network, as a^2 indicated that the load increase was compensated by generator 2, the cheapest and optimal one.

Therefore a^1 is more dangerous than a^2, even though they have the same sparsity, which motivates the need for tools to analyze the system vulnerability while evaluating the attack impact.

4.3.4 Impact Aware Security Metric

In this section we propose a novel impact aware security metric addressing the following problem:

Problem 2. *Given a data attack targeting measurement k, what is the maximum impact to the network operation by undetectable attacks with a given sparsity?*

A solution to this problem is obtained by first constraining the attack sparsity and then finding the maximum impact to the power network operation, resulting in the proposed metric. Problem 2 is addressed for a given initial demand P^d and the corresponding optimal dispatch P^{g*} using the results in the previous sections. Our goal is to provide to the operator impact aware security metrics designed to quantify the vulnerability to *and* impact of data attacks on the several measurements. As a result, the measurements with the highest index values are likely candidates for protection, similarly to the approach in (Dán and Sandberg, 2010) and mentioned in Chapter 3.

A solution to Problem 2 is now addressed, considering the consequence to the operator to be the real re-dispatch profit $\mathcal{P}_a$. Recalling that $H_{DC}\theta$ is a vector of all possible measurements, since $H_{DC} \in \mathbb{R}^{N+N_b\times N-1}$, consider the measurement model $z = \Gamma H_{DC}\theta$ with $\Gamma \in \mathbb{R}^{m\times m}$ being a diagonal matrix of binary diagonal entries indicating whether a particular variable $[H_{DC}\theta]_i$ is measured (Γ_{ii}=1) or not (Γ_{ii}=0). Constraining the attack effort with $\|\Gamma a\|_0 \leq C$, Problem 2 then corresponds to the following optimization problem

$$\sigma_k = \max_{c} |\mathcal{P}_a| \tag{4.24}$$

$$a = H_{DC}c \tag{4.25}$$

$$a = \begin{bmatrix} (C_g a_g - a_d)^\top & a_f^\top \end{bmatrix}^\top \tag{4.26}$$

$$\hat{\xi} \leq \hat{\mathcal{P}}_a \tag{4.27}$$

$$C \geq \|\Gamma a\|_0 \tag{4.28}$$

$$\epsilon = |e_k^\top a|. \tag{4.29}$$

Remark 4.3.5. *Another possible approach is to modify the security metric in (3.17) by further constraining the attacks so that the resulting biased estimate leads the operator to apply the corrupted DC-OPF recommendation and the power network operation is indeed affected. These could be posed as additional constraints formulated in terms of $\hat{\mathcal{P}}_a$ and $\mathcal{P}_a$, respectively.*

Computation issues

The feasible set of (4.24) is not convex due to the equality constraint (4.29) and the inequality (4.28). Regarding the inequality, a possible relaxation is to instead consider the inequality $\|\Gamma a\|_1 \leq C$, since the 1-norm is known to provide reasonable approximations of sparsity, as mentioned in (Sou et al., 2011). Even with this relaxation, we would still have the feasible set as the union of two convex sets due to (4.29). Furthermore, the objective function is not concave. However, given the linear approximation of $\mathcal{P}_a$ in (4.20), for each convex feasible set one could solve two Linear Programming (LP) problems, namely min $\mathcal{P}_a$ and max $\mathcal{P}_a$, and compare the respective solutions. Thus the security index σ_k can be computed by solving and comparing the solutions of four LP problems.

An alternative method to compute the solution to (4.24) is as follows. Suppose that, for a given sparsity level C, all the possible C-sparse attacks are obtained. This could be achieved, for instance, using the approaches in (Dán and Sandberg, 2010) and (Giani et al., 2011). Then the optimization problem (4.24) is reduced to an iterative procedure over all the $C-$sparse attacks, evaluating $|\mathcal{P}_a|$ of each attack satisfying the constraints. This approach is illustrated in the following numerical example.

Numerical example

An impact-aware security metric considering the data attack impact is now computed for the VIKING 40−bus benchmark using the DC network model. In this example, we considered that all active power flows and injections measurements are taken, i.e. there is full measurement redundancy. Note that the impact-aware security metric presented here is not σ_k as in (4.24), but rather an approximation, $\tilde{\sigma}_k$, computed using an iterative search over the sparse attack vectors identified.

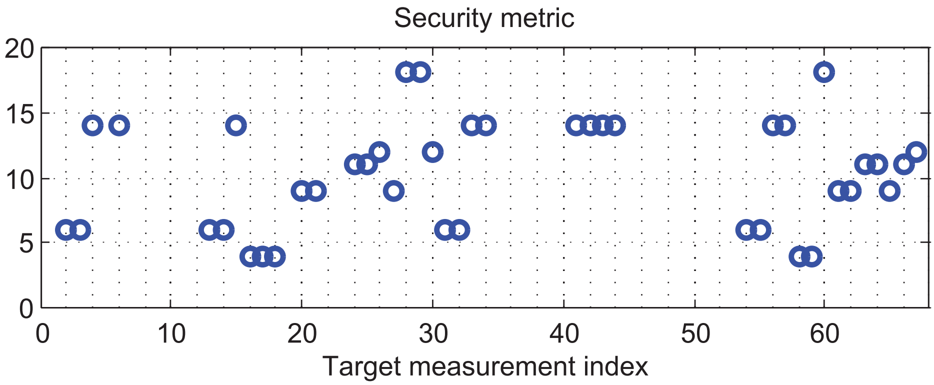


Figure 4.2: The security metric ρ_k for each measurement k, computed assuming all measurements are available.

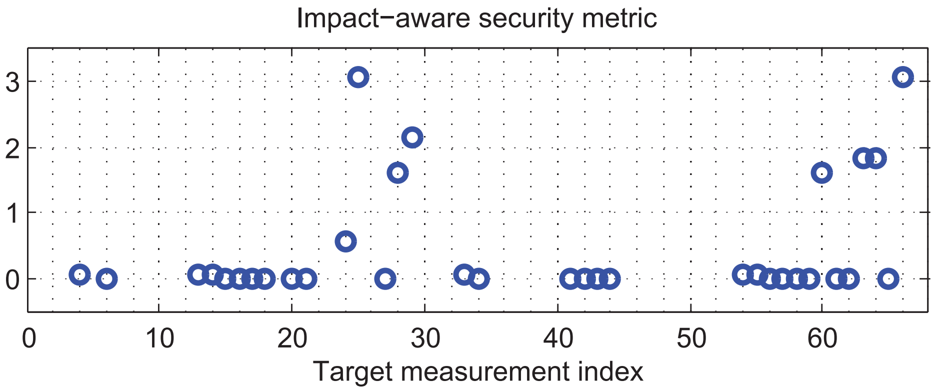


Figure 4.3: An impact-aware security metric $\tilde{\sigma}_k$, corresponding to the percentage increase in the transmission losses due to stealthy deception attack on each measurement k.

Following the algorithms in (Dán and Sandberg, 2010), a set of the sparsest attacks targeting each measurement k were identified, with $a_k = 1$pu. The security metric without considering the attack impact, ρ_k proposed in Chapter 3, is shown in Figure 4.2.

Regarding the attack impact, we considered the DC-OPF objective function as the total active losses $c(P^g, P^d) = \sum_{i=1}^{N_b} r_i (P_i^f)^2 = P^{f\top} D_r P^f$, where r_i is the resistance of the transmission line i and $D_r = \operatorname{diag}(r_1, \ldots, r_{N_b})$. Note, however, that there are other relevant objective functions that can also be included in the DC-OPF, such as the generator costs. The nominal DC-OPF solution was considered to have no saturated lines. Using the iterative methodology described in Section 4.3.4, we computed the economic impact $\mathcal{P}_a$ for each of the sparse attacks previously identified. The relative increase in the total losses due to the data attack on each measurement is shown in Figure 4.3.

Comparing the results in Figures 4.2 and 4.3, we observe that most sparse attacks have a low economic impact on the power network operation. In fact, the attacks for which the economic impact is considerable corrupt more than 10 measurements. Therefore it is important to include impact analysis when assessing the vulnerability of the system and deploying protective solutions.

Enforcing Assumption 4.3.2

The impact analysis in this section remains valid as long as Assumption 4.3.2 holds. To strengthen the validity of the analysis, necessary and sufficient conditions for Assumption 4.3.2 to hold are now provided, which may be included as additional constraints when computing the impact aware security metric (4.24).

Theorem 4.3.3. *The necessary and sufficient conditions for Assumption 4.3.2 to hold are*

$$\begin{bmatrix} -H_1 T_\lambda \\ H_0(F_g T_g + F_d) \end{bmatrix} a_d < \begin{bmatrix} H_1 \lambda^* \\ H_0(-F_g P^{g*} - F_d P^d - F_0) \end{bmatrix}$$
$$\begin{bmatrix} H_0 T_\lambda \\ H_1(F_g T_g + F_d) \end{bmatrix} a_d = \begin{bmatrix} 0 \\ H_1(-F_g P^{g*} - F_d P^d - F_0) \end{bmatrix}.$$

Proof. Consider the set of primal and dual variables computed based on corrupted data using (4.8) and (4.13)

$$\begin{bmatrix} \hat{P}_a^{g*} \\ \hat{\lambda}_a^* \\ \hat{\nu}_a^* \end{bmatrix} = \begin{bmatrix} T_g \\ T_\lambda \\ T_\nu \end{bmatrix} a_d + \begin{bmatrix} P^{g*} \\ \lambda^* \\ \nu^* \end{bmatrix}.$$

Note that Assumption 4.3.2 is equivalent to the optimality of the primal and dual variables $\begin{bmatrix} \hat{P}_a^{g*\top} & \hat{\lambda}_a^{*\top} & \hat{\nu}_a^{*\top} \end{bmatrix}^\top$ for the DC-OPF problem with corrupted data. Thus the necessary and sufficient conditions for Assumption 4.3.2 to hold correspond to the KKT optimality conditions.

By construction, all primal and dual variables computed using (4.8) satisfy $\nabla L(\hat{P}_a^{g*}, \hat{\nu}_a^*, \hat{\lambda}_a^*) = 0$ and $h(\hat{P}_a^{g*}, \hat{P}_a^d) = 0$. Thus only the inequality constraints need to be considered.

Regarding the dual variables of the inequality constraints, the variables corresponding to active constraints are positive, while the remaining variables are zero, yielding

$$\begin{aligned} H_1\hat{\lambda}_a^* &= H_1 T_\lambda a_d + H_1\lambda^* > 0 \\ H_0\hat{\lambda}_a^* &= H_0 T_\lambda a_d = 0. \end{aligned}$$

To conclude the proof, note that the primal variables also need to be constrained so that the active and inactive inequality constraints remain unchanged, leading to the following conditions

$$\begin{aligned} H_1(F_g T_g + F_d)a_d &= H_1(-F_g P^{g*} - F_d P^d - F_0) \\ H_0(F_g T_g + F_d)a_d &< H_0(-F_g P^{g*} - F_d P^d - F_0). \end{aligned}$$

□

4.4 Summary

This chapter addressed the DC-OPF based power network operation in the presence of stealthy data corruption attacks on the measurements. Given the biased estimates resulting from the measurement corruption, we derived analytical expressions characterizing the behavior of the DC-OPF for attacks that do not affect the system operating constraints. Based on these expressions, we discussed under what conditions a human operator would have the incentive to close the loop over the corrupted measurements, thus applying erroneous actions to the network. The economic impact of applying these erroneous control actions was also discussed and analytically characterized.

As an important outcome of this study, an impact-aware security metric for the measurements is proposed, quantifying their vulnerability to stealthy attacks and impact on the power network operation. The proposed metric is illustrated through numerical examples using the VIKING 40−bus testbed. The numerical results indicate that a large number of the possible stealthy attacks, especially the most sparse ones, do not appear to be dangerous, in the sense that they yield a small impact on the power network operation cost. Note that this observation depends on the power network physical and operation cost models.

Chapter 5

Distributed Fault Diagnosis in Networked Systems

Automatic detection of system faults is of growing importance as the size and complexity of systems rapidly increase. Most of the available literature on model-based fault detection and isolation (FDI) focuses on centralized systems where the FDI scheme has access to all the available measurements and the objective is to detect and isolate faults occurring in any part of the system (Chen and Patton, 1999; Ding, 2008; Isermann, 2004). Distributed implementations are more suitable than centralized for large-scale interconnected dynamical systems such as power networks and multi-agent systems due to its lower complexity and less use of network resources (Siljak, 1991). Traditional FDI schemes may not be applied to distributed systems, since not all measurements are available in every node.

The design of distributed FDI for large-scale networked systems is addressed in this chapter. The outline is as follows. The problem of distributed FDI for a class of networked systems is formulated in Section 5.2. In Section 5.3 we recall the existing FDI tools and new distributed solutions are proposed in Section 5.4. The complexity of the proposed solution and methods to reduce it are discussed in Section 5.5. In Section 5.6 the application of the results to two practical problems is studied. A summary of the chapter is given in Section 5.7.

5.1 Contributions and Related Work

Some recent work has been done on the design of distributed FDI scheme. In (Ding et al., 2008), a bank of decentralized observers is built where each observer contains the model of the entire system and receives both measurements from the local subsystem and information transmitted from other observers. A similar approach is taken in (Chung et al., 2001) where the observers communicate with each other, but they only possess models of their respective local subsystems. A mixing procedure is used to reconstruct the state of the overall system from the local estimates. Recently a distributed FDI scheme for a network of interconnected first-order systems was proposed. The authors analyzed limitations on fault detectability and isolability in a system-theoretic perspective (Pasqualetti et al., 2010).

Power networks are large-scale spatially distributed systems. Being a critical infrastructure, they possess strict safety and reliability constraints (Shahidehpour et al., 2005b). Monitoring the state of the system is essential to guarantee safety. Currently this is typically done in a centralized control center through a single state estimator. The core methodology for state estimation of power systems dates from 1970, (Abur and Exposito, 2004; Schweppe and Wildes, 1970). Due to the low sampling frequency of the sensors in these systems a steady-state approach is taken, which only allow for an over-constrained operation of the system to ensure reliability. Furthermore dynamic faults such as generator electro-mechanical oscillations may pass undetected by schemes based on steady-state models and measurements, possibly leading to cascade failures.

In recent years, measurement units with higher sampling rate have been developed, e.g., Phasor Measurement Units (PMU), opening the way to dynamic state estimators and observer-based fault detection schemes taking in account the dynamics of the system. Such centralized FDI schemes have been proposed in the recent literature, see (Aldeen and Crusca, 2006; Demetriou, 2005; Scholtz and Lesieutre, 2008). However, to the best of our knowledge, no distributed method has been proposed to carry out FDI in power networks, despite their inherent decentralized nature.

In this chapter we address the problem of distributed FDI in a network of nodes with double integrator dynamics, whose interactions are described by a distributed control law. We show how FDI for some power networks and distributed robotic systems fit the problem description. We design continuous-time unknown input observers (UIOs) to achieve the goal. The existence of such observers is established for various conditions on the node interactions. The results are illustrated on examples in power networks and autonomous mobile node formations.

5.2 Problem Formulation

Consider a network of N interconnected nodes and let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the underlying graph, where $\mathcal{V} \triangleq \{i\}_1^N$ is the vertex set, with $i \in \mathcal{V}$ corresponding to node i, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. The undirected edge $\{i, j\}$ is incident to vertices i and j if nodes i and j share a communication link. Moreover, $\mathcal{N}_i = \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\}$ is the neighborhood set of i. Each node i is assumed to have double integrator dynamics

$$\dot{\xi}_i(t) = \zeta_i(t) \tag{5.1a}$$

$$\dot{\zeta}_i(t) = u_i(t) + v_i(t), \tag{5.1b}$$

where v_i is a scalar known external input, ξ_i, ζ_i are the scalar states, and u_i is the control given by the linear control law

$$u_i(t) = -\kappa_i\zeta_i(t) + \sum_{j\in\mathcal{N}_i} w_{ij}\left[(\xi_j(t) - \xi_i(t)) + \gamma(\zeta_j(t) - \zeta_i(t))\right], \tag{5.2}$$

where $w_{ij} > 0$ is the edge weight and $\kappa_i, \gamma \geq 0$ for $i, j = 1, \ldots, N$. We say that node $k \in \mathcal{V}$ is faulty if for some functions $f_{\xi k}(t)$ and $f_{\zeta k}(t)$ not identical to zero either $\dot{\xi}_k(t) = \zeta_k(t) + f_{\xi k}(t)$, or $\dot{\zeta}_k(t) = u_k(t) + v_k(t) + f_{\zeta k}(t)$. The functions $f_{\xi k}(t)$ and $f_{\zeta k}(t)$ are denoted fault signals. It is assumed that the faulty node injects a fault in only one of the states.

Remark 5.2.1. *The variables ξ_i and ζ_i can be interpreted as position and velocity of node i, respectively, for a mobile system, or as phase and frequency in the context of power networks, as further discussed in Section 5.6.*

The closed-loop dynamics of the networked system in the presence of faults can be written as

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bv(t) + B_f f(t)\\ y(t) &= Cx(t),\end{aligned} \tag{5.3}$$

where $x(t) = [\xi_1(t), \ldots,\ \xi_N(t),\ \zeta_1(t), \ldots,\ \zeta_N(t)]^\top$. The signal $f(t) \in \mathbb{R}^m$ is a vector of unknown fault signals, $y(t) \in \mathbb{R}^p$ is the output vector, and A, B, B_f, and C are matrices of appropriate dimensions. More specifically, we have

$$A = \begin{bmatrix} 0_N & I_N \\ -\mathcal{L} & -\gamma\mathcal{L} - \kappa I_N \end{bmatrix},\ B = \begin{bmatrix} 0_N \\ I_N \end{bmatrix}, \tag{5.4}$$

where $\mathcal{L}$ is a Laplacian matrix and $\kappa = \text{diag}(\kappa_1, \ldots, \kappa_N)$. The ij-th entry of $\mathcal{L}$, $\mathcal{L}_{ij}$, is equal to -1 if i and j share a link and zero otherwise, moreover, $\mathcal{L}_{ii} = -\sum_{j\in\mathcal{N}_i} \mathcal{L}_{ij}$. We call the faults $f(t)$ additive faults, see (Ding, 2008).

Before stating the problem, we define what is meant by fault *detectability* and *isolability* for (5.3) in the following (Ding, 2008).

Definition 5.2.1 (Detectable and Isolable Fault)**.** *Given the system (5.3), m scalar faults $f(t) = [f_1(t), \ldots, f_m(t)]^\top$ are detectable and isolable if*

$$\textit{rank}\begin{bmatrix} sI - A & B_f \\ C & 0 \end{bmatrix} = n + m$$

for almost all $s \in \mathbb{C}$.

A fault is thus detectable if the transfer function from $f_k(t)$ to $y(t)$ is not identical to zero. Isolable faults relate to input observability and means that any simultaneous occurrence of faults should lead to a change in the output. We further note that the FDI scheme proposed in this paper can detect almost all faults. That is, there may be values $s \in \mathbb{C}$ for which the matrix in Definition 5.2.1 does not have full rank. Hence there may be some faults generating zero dynamics, which, by definition, do not appear in the system output. These faults cannot be detected using the scheme proposed in this paper and may be seen as the dynamic equivalent to the stealthy attacks discussed in Chapter 3.

Note that B_f is a matrix such that each of its columns b_{f_k} has its entries corresponding to the states of node k as the only non-zero entries. Each node k has a scalar fault signal $f_k(t)$ with distribution vector b_{f_k}. We say node k is faulty if $f_k(t)$ is not identical to zero.

The measurement matrix C may be viewed as a design parameter to be chosen in order to ensure the feasibility of the distributed FDI scheme with respect to a predetermined set of faults to be detected. We assume that each node i only measures states within its neighborhood, thus ensuring the distributed nature of the FDI scheme. As it will be shown later, the specific structure of a feasible local measurement matrix will depend on the faults to be detected.

In this paper, we solve the following problems:

Problem 3. *How can each node of the network detect and isolate a faulty agent?*

Problem 4. *How can the faulty agent be automatically removed?*

We propose a solution to these two problems for two different classes of distributed control laws in the coming sections. In next section we introduce the mathematical tool that we use. Then, in Section 5.4 we solve Problems 3 and 4, and give conditions for when the solutions exist.

5.3 Model-Based Fault Detection Preliminaries

We now present UIOs and their application to FDI for centralized linear control systems (Chen and Patton, 1999; Ding, 2008). A common technique used in model-based fault diagnosis is to generate a set of residuals which indicate the presence of a fault. The residual is a fault indicator computed from the difference between the measurements and their estimates. It should be close to zero if and only if the fault is not present.

Consider the linear fault-free system under the influence of an unknown input $d(t) \in \mathbb{R}^{m-1}$ described by

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bv(t) + Ed(t) \\ y(t) &= Cx(t). \end{aligned} \tag{5.5}$$

The system in presence of faults is given by

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bv(t) + Ed(t) + B_f f(t) \\ y(t) &= Cx(t). \end{aligned} \tag{5.6}$$

We assume that the matrices E and B_f have full column rank.

Remark 5.3.1. *Note that the condition on B_f being full column rank is not restrictive, since any singular matrix $D \in \mathbb{R}^{n \times l}$ can be decomposed in $D = D_1 D_2$, with D_1 having full column rank. This implies, however, that not all faults are isolable, as follows from the analysis in Section 5.4.*

The matrix E is called a disturbance distribution matrix, since it contains information on how a vector of unknown input disturbances affect the states of the system.

A full-order observer for the fault-free system (5.5) is described by:

$$\begin{aligned} \dot{z}(t) &= Fz(t) + TBv(t) + Ky(t) \\ \hat{x}(t) &= z(t) + Hy(t), \end{aligned} \tag{5.7}$$

where $\hat{x}(t) \in \mathbb{R}^n$ is the estimated state and $z(t) \in \mathbb{R}^n$ is the observer's state. Note that if we choose $F = A - KC$, $T = I$, and $H = 0$ we have a full-order Luenberger observer. The observer matrices must be designed to achieve the decoupling from the unknown input and meet requirements on the stability of the observer. Choosing the matrices F, T, K, H to satisfy the following conditions

$$\begin{aligned} &F = (A - HCA - K_1 C), \quad T = (I - HC) \\ &K = K_1 + K_2, \quad K_2 = FH, \quad (HC - I)E = 0, \end{aligned} \tag{5.8}$$

we have the estimation error dynamics

$$\dot{e}(t) = Fe(t). \tag{5.9}$$

where $e(t) = x(t) - \hat{x}(t)$. Now we have the following definition of a UIO.

Definition 5.3.1 (UIO)**.** *A state observer is a UIO if the state estimation error $e(t)$ approaches zero asymptotically, regardless of the presence of the unknown input $d(t)$.*

We conclude that if (5.8) is satisfied and F is stable, then the observer (5.7) is a UIO. The following proposition from (Chen and Patton, 1999) formalizes this.

Proposition 5.3.2. *There exists a UIO for* (5.5) *if and only if*

1. *rank*$(CE) =$ *rank*(E)
2. $(C, A - HCA)$ *is a detectable pair, where H is given by* (5.8)*.*

For a proof and more details the reader is referred to (Chen and Patton, 1999; Ding, 2008). As suggested in (Chen and Patton, 1999), a possible method of detecting and isolating the faults is to use the so called generalized observer scheme (GOS), where we construct a bank of observers generating a structured set of residuals such that each residual is decoupled from one and only one fault, but being sensitive to all other faults. Suppose there is a single fault, $f_i(t) \neq 0$. In order to render the observer insensitive to $f_i(t)$, this fault is regarded as an unknown input. The system (5.6) for $d \equiv 0$ is equal to

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bv(t) + B_{f_{-i}} f_{-i}(t) + b_{f_i} f_i(t) \\ y(t) &= Cx(t), \end{aligned} \tag{5.10}$$

where b_{f_i} is the i-th column of B_f, $f_i(t)$ the i-th component of $f(t)$, $B_{f_{-i}}$ is B_f with the i-th column deleted and $f_{-i}(t)$ the fault vector $f(t)$ with its i-th component removed. Note that $f_i(t)$ can be considered as a disturbance that we want to decouple (b_{f_i} is analogous to E in (5.6)).The UIO decoupled from b_{f_i} has thus the same structure as (5.7) and is described by

$$\begin{aligned} \dot{z}_i(t) &= F_i z_i(t) + T_i B v(t) + K_i y(t) \\ \hat{x}_i(t) &= z_i(t) + H_i y(t). \end{aligned} . \tag{5.11}$$

We introduce residuals to indicated faults.

Definition 5.3.3. *A residual $r_i(t)$ is a fault indicator function that satisfies*

$$\|r_i(t)\| = 0 \Leftrightarrow \|f_{-i}(t)\| = 0.$$

It is easy to show that we have the following observer error and residual dynamics

$$\begin{aligned} \dot{e}_i(t) &= F_i e_i(t) - T_i B_{f_{-i}} f_{-i}(t) \\ r_i(t) &= C e_i(t) \end{aligned} . \tag{5.12}$$

where $e_i(t) = x(t) - \hat{x}_i(t)$ is the observer error and $r_i(t)$ is the corresponding residual. Note that the residual dynamics are driven by the k-th fault if $T_i b_{f_k} \neq 0, k \neq i$.

We introduce the following detection and isolation condition for fault $f_i(t)$,

$$\begin{aligned} &\|r_i(t)\| < \Theta_{f_i} \\ &\|r_j(t)\| \geq \Theta_{f_j} \ , \forall j \neq i, \end{aligned} \tag{5.13}$$

where $\Theta_{f_i}, \Theta_{f_j} > 0$ are isolation thresholds, which can be constant or time varying. If (5.13) is satisfied, we conclude that there is a fault affecting the i-th component of the system. Note that the selection of Θ_{f_i} is particularly important. The interested reader may refer to (Frank and Ding, 1997) and references there-in for more information

The approach presented above is feasible only if a single additive fault is present. To isolate multiple faults, one can repeat the abovementioned procedure for each of the potential fault combinations. We can derive similar observers for all faults and then use (5.13) to isolate each of them. Next we show that one can construct UIOs also for classes of networked systems.

5.4 Distributed FDI for Networked Systems

In Sections 5.4.1 and 5.4.2 we solve Problem 3 of Section 5.2 by considering two different distributed control laws that are special cases of (5.2) and show that UIOs can under certain conditions be applied in both cases. Section 5.4.3 presents the solution to Problem 4.

5.4.1 UIO for Position Distributed Control

Consider the networked system introduced in Section 5.2 with the following control law

$$m_i u_i(t) = -d_i \zeta_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij} \left(\xi_j(t) - \xi_i(t) \right). \tag{5.14}$$

where $m_i, w_{ij}, d_i > 0$. If we make the physical interpretation that $\xi_i(t)$ and $\zeta_i(t)$ are position and velocity of node i, m_i can be interpreted as the agent's mass. The nodes under the control law (5.14) move towards the position of their neighbors while damping their current velocity.

As in Section 5.2, assume that

$$\dot{\xi}_k(t) = \zeta_k(t) + f_k(t) \tag{5.15}$$

where $f_k(t)$ corresponds to a fault in node k. In the presence of this fault, we have

$$\dot{x}(t) = Ax(t) + b_f^k f_k(t) \tag{5.16}$$

where

$$\begin{aligned} A &= \begin{bmatrix} 0_N & I_N \\ -\bar{M}\mathcal{L} & -\bar{M}\bar{D} \end{bmatrix} \\ B &= \begin{bmatrix} 0_N \ \bar{M} \end{bmatrix}^\top \\ \bar{M} &= \operatorname{diag}\left(\frac{1}{m_1}, \cdots, \frac{1}{m_N} \right) \\ \bar{D} &= \operatorname{diag}\left(d_1, \cdots, d_N \right) \\ b_f^k &= [\bar{b}_f^{k\top} \ 0_{1\times N}]^\top, \end{aligned} \tag{5.17}$$

with $\bar{b}_f^k$ being an N dimensional vector with all zero entries except one that corresponds to the faulty node k. Furthermore, we assume the nodes have access to

$$y_i(t) = C_i x(t), \quad C_i = \begin{bmatrix} \bar{C}_i & 0_{|\tilde{\mathcal{N}}_i| \times N} \end{bmatrix}, \quad i = 1, \ldots, N, \tag{5.18}$$

with $\bar{C}_i$ being an $|\tilde{\mathcal{N}}_i|$ by N matrix with full row rank, where each of the rows have all zero entries except for one entry at the j-th position that corresponds to those nodes that are neighbors of i, where $\tilde{\mathcal{N}}_i = \mathcal{N}_i \cup \{i\}$ and $j \in \tilde{\mathcal{N}}_i$.

To solve Problem 3, we show that one can construct a UIO at any given node i under the control law (5.14) using measurements (5.18).

Theorem 5.4.1. *Consider the distributed control system with a fault in node k given by (5.16) and local measurments (5.18). If $\mathcal{G}$ is connected and $k \in \mathcal{N}_i$, then there exists a UIO for node i.*

Proof. First we show that

$$\operatorname{rank}\left(C_i b_f^k\right) = \operatorname{rank}\left(b_f^k\right) = 1.$$

Denote the row of C_i that reads the output of node k, c_i^k. It is obvious that $c_i^k b_f^k = 1$ and $c_i^j b_f^k = 0,\ j \neq k$. Hence, $C_i b_f^k$ is a vector with zero entries except one which is equal to 1, thus the rank is equal to 1. This condition is equivalent to condition (1) of Proposition 5.3.2.

Then we show that $\operatorname{rank}(\mathcal{D}) = 2N + 1$ for all $\operatorname{Re}(s) \geq 0$ where

$$\mathcal{D} = \begin{bmatrix} sI_{2N} - A & b_f^k \\ C_i & 0_{|\tilde{\mathcal{N}}_i| \times 1} \end{bmatrix},$$

which is equivalent to to Proposition 5.3.2 (2) and also shows the fault is detectable according to Definition 5.2.1. We have

$$\operatorname{rank}(\mathcal{D}) = \operatorname{rank} \begin{bmatrix} sI_N & -I_N & \bar{b}_f^k \\ \bar{M}\mathcal{L} & sI_N + \bar{D}\bar{M} & 0_{N \times 1} \\ \bar{C}_i & 0_{|\tilde{\mathcal{N}}_i| \times N} & 0_{|\tilde{\mathcal{N}}_i| \times 1} \end{bmatrix}$$

Applying some row and column operations we obtain

$$\operatorname{rank}(\mathcal{D}) = \operatorname{rank} \begin{bmatrix} 0_N & -I_N & \bar{b}_f^k \\ a(s) & 0_N & b(s) \\ \bar{C}_i & 0_{|\tilde{\mathcal{N}}_i| \times N} & 0_{|\tilde{\mathcal{N}}_i| \times 1} \end{bmatrix},$$

with $a(s) = s^2 I_N + s\bar{D}\bar{M} + \bar{M}\mathcal{L}$, and $b(s) = (sI_N + \bar{D}\bar{M})\bar{b}_f^k$.

We apply a state transformation

$$\begin{aligned} \bar{x} = Px = [&\xi_{\tilde{i}_1}, \cdots, \xi_{\tilde{i}_{|\tilde{\mathcal{N}}_i|}}, \xi_{\bar{i}_1}, \cdots, \xi_{\bar{i}_{|\bar{N}_i|}}, \\ &\zeta_{\tilde{i}_1}, \cdots, \zeta_{\tilde{i}_{|\tilde{\mathcal{N}}_i|}}, \zeta_{\bar{i}_1}, \cdots, \zeta_{\bar{i}_{|\bar{N}_i|}}]^\top, \end{aligned}$$

where $\tilde{i}_j \in \tilde{\mathcal{N}}_i$, $\bar{i}_j \in \bar{N}_i$, and $\bar{C}_i^* = \bar{C}_i P = [I_{|\tilde{\mathcal{N}}_i|} 0_{|\tilde{\mathcal{N}}_i| \times \bar{N}_i}]$, where $\tilde{\mathcal{N}}_i = i \cup \mathcal{N}_i$ and $\bar{N}_i = \mathcal{V} \setminus \tilde{\mathcal{N}}_i$. After this operation we can write the Laplacian as $\bar{\mathcal{L}} = P^{-1}\mathcal{L}P = \begin{bmatrix} \mathcal{L}_{|\tilde{\mathcal{N}}_i|} & l_{|\tilde{\mathcal{N}}_i| \times |\bar{N}_i|} \\ l_{|\bar{N}_i| \times |\tilde{\mathcal{N}}_i|} & \mathcal{L}_{|\bar{N}_i}| \end{bmatrix}$.

Furthermore we have $\tilde{b}_f^k = P^{-1}\bar{b}_f^k$, $\tilde{b}_f^{k*} = P^{-1}(sI_N + \bar{D}\bar{M})\bar{b}_f^k$, $P^{-1}\bar{M}P = \begin{bmatrix} \bar{M}_{1|\tilde{\mathcal{N}}_i|} & 0_{|\tilde{\mathcal{N}}_i|\times|\bar{N}_i|} \\ 0_{|\bar{N}_i|\times|\tilde{\mathcal{N}}_i|} & \bar{M}_{2|\bar{N}_i}| \end{bmatrix}$, and $P^{-1}\bar{D}P = \begin{bmatrix} \bar{D}_{1|\tilde{\mathcal{N}}_i|} & 0_{|\tilde{\mathcal{N}}_i|\times|\bar{N}_i|} \\ 0_{|\bar{N}_i|\times|\tilde{\mathcal{N}}_i|} & \bar{D}_{2|\bar{N}_i}| \end{bmatrix}$.

After applying the transformation we have

$$\operatorname{rank}(\mathcal{D}) = \operatorname{rank}\begin{bmatrix} 0_{|\bar{N}|\times|\tilde{\mathcal{N}}_i|} & 0_{|\bar{N}_i|\times|\bar{N}_i|} & -I_N & \tilde{b}_f^k \\ c(s) & \bar{M}_1 l_{|\tilde{\mathcal{N}}_i|\times|\bar{N}_i|} & 0_{|\tilde{\mathcal{N}}_i|\times N} & \tilde{b}_f^{k*} \\ \bar{M}_2 l_{|\bar{N}_i|\times|\tilde{\mathcal{N}}_i|} & d(s) & 0_{|\bar{N}_i|\times N} & 0_{|\bar{N}_i|\times 1} \\ I_{|\tilde{\mathcal{N}}_i|} & 0_{|\tilde{\mathcal{N}}_i|\times|\bar{N}_i|} & 0_{|\tilde{\mathcal{N}}_i|\times N} & 0_{|\tilde{\mathcal{N}}_i|\times 1} \end{bmatrix},$$

with $c(s) = \bar{M}_1\mathcal{L}_{|\tilde{\mathcal{N}}_i|} + s^2 I_{|\tilde{\mathcal{N}}_i|} + s\bar{M}_1\bar{D}_1$, and $d(s) = \bar{M}_2\mathcal{L}_{|\bar{N}_i|} + s^2 I_{|\bar{N}_i|} + s\bar{M}_2\bar{D}_2$. It is evident that the first and the third columns are independent of the rest, thus

$$\operatorname{rank}(\mathcal{D}) = |\tilde{\mathcal{N}}_i| + N + \operatorname{rank}\begin{bmatrix} \bar{M}_1 l_{|\tilde{\mathcal{N}}_i|\times|\bar{N}_i|} & \tilde{b}_f^{k*} \\ \bar{M}_2\mathcal{L}_{|\bar{N}_i|} + s^2 I_{|\bar{N}_i|} + s\bar{M}_2\bar{D}_2 & 0_{|\bar{N}_i|\times 1} \end{bmatrix}.$$

We know from (Barooah and Hespanha, 2007) that any principal submatrix of the Laplacian matrix is invertible so the last column is independent of the rest as well, hence $\operatorname{rank}(\mathcal{D}) = |\tilde{\mathcal{N}}_i| + N + |\bar{N}_i| + 1 = 2N + 1$. This rank equality is equivalent to condition (2) of Proposition 5.3.2 (Chen and Patton, 1999). Satisfying the two conditions of Proposition 5.3.2 the existence of a UIO for the system (5.16) with measurements (5.18) and a fault in node k is established. □

Remark 5.4.1. *Note that if the graph is not connected, the networked system* (5.16) *can be decomposed into several decoupled subsystems, each corresponding to a connected subset of the network. The conclusion of Theorem 5.4.1 then applies to each subsystem.*

The existence of a UIO according to Theorem 5.4.1 leads to the possibility to detect a fault at node k from a neighboring node i using the methods described in Section 5.3.

In Theorem 5.4.1 we stated that a fault in ξ_k can be isolated with the measurements of the form (5.18). In the next theorem we identify faults that cannot be isolated.

Theorem 5.4.2. *Consider the system* (5.16). *For any of the following pairs of* C_i *and* b_f^k, *no UIO of the form* (5.7) *exists:*

(i) $b_f^k = [\bar{b}_f^{k\top}\ 0_{1\times N}]^\top$, $C_i = \begin{bmatrix} 0_{|\tilde{\mathcal{N}}_i|\times N} & \bar{C}_i \end{bmatrix}$

(ii) $b_f^k = [0_{1\times N}\ \bar{b}_f^{k\top}]^\top$, $C_i = \begin{bmatrix} 0_{|\tilde{\mathcal{N}}_i|\times N} & \bar{C}_i \end{bmatrix}$

(iii) $b_f^k = [0_{1\times N}\ \bar{b}_f^{k\top}]^\top$, $C_i = \begin{bmatrix} \bar{C}_i & 0_{|\tilde{\mathcal{N}}_i|\times N} \end{bmatrix}$

Proof. To see that no UIO exists for (i) and (iii), we simply verify that

$$\operatorname{rank}\left(C_i b_f^k\right) = \operatorname{rank}\left(b_f^k\right) = 0,$$

so the first condition of Proposition 5.3.2 is not satisfied. For (ii), similar to the calculations in proof of Theorem 5.4.1, for the case where $s = 0$, we have

$$\operatorname{rank}(\mathcal{D}) = \operatorname{rank}\begin{bmatrix} 0_N & -I_N & \bar{b}_f^k \\ \bar{M}\mathcal{L} & 0_N & \bar{D}\bar{M}\bar{b}_f^k \\ 0_{|\tilde{\mathcal{N}}_i| \times N} & \bar{C}_i & 0_{|\tilde{\mathcal{N}}_i| \times 1} \end{bmatrix}. \tag{5.19}$$

Recall that $\mathcal{L}$ is rank deficient. Then, it follows that the first column block above is not full column rank. Hence the second condition of Proposition 5.3.2 is not satisfied. □

Cases (i) and (iii) of Theorem 5.4.2 suggest that if there is an unknown input affecting one of the states of one of the nodes in a network, it is not possible to have a UIO without measuring the same state throughout the network as the one affected by the unknown input. For example, if a fault is affecting the velocity of one of the nodes, by measuring positions alone we cannot have a UIO to observe the states of the network. On the other hand, in Case (ii) we see that the first condition of Proposition 5.3.2 is satisfied, but a UIO still does not exist. What happens in this case is that the system is not detectable, as seen by observing the first two columns of (5.19). However, by having access to more measurements one can construct a UIO to detect and isolate faults as seen next.

We now introduce conditions for existence of a UIO to detect the fault

$$\dot{\zeta}_k(t) = u_i(t) + v_i(t) + f_k(t), \tag{5.20}$$

where again $f_k(t)$ corresponds to a fault in node k.

Theorem 5.4.3. *Consider the distributed control system with a fault in node k given by (5.16) and local measurments (5.18) with $C_i = \begin{bmatrix} \bar{C}_i & 0_{|\tilde{\mathcal{N}}_i| \times N} \\ 0_{|\tilde{\mathcal{N}}_i| \times N} & \bar{C}_i \end{bmatrix}$, where $\bar{C}_i$ is a $|\tilde{\mathcal{N}}_i|$ by N matrix, and $b_f^{k\top} = \begin{bmatrix} 0_{1\times N} & \bar{b}_f^{k\top} \end{bmatrix}$ with b_f^k being an N by 1 vector with k-th entry as its only nonzero entry. If $\mathcal{G}$ is connected and $k \in \mathcal{N}_i$, then there exists a UIO for node i.*

5.4.2 UIO for Position–Velocity Distributed Control

Now we consider the existence of UIOs for the distributed control law:

$$u_i(t) = \sum_{j \in \mathcal{N}_i} w_{ij} \left[(\xi_j(t) - \xi_i(t)) + \gamma(\zeta_j(t) - \zeta_i(t)) \right]. \tag{5.21}$$

Again, interpreting $\xi_i(t)$ and $\zeta_i(t)$ to be position and velocity of node i, the nodes under the control law described by (5.21) move towards the position of their neighbors while penalizing not only the position differences (as previously) but also penalizing the velocity difference. The dynamics of the networked system with a faulty node k is

$$\dot{x}(t) = Ax(t) + b_f^k f_k(t) \tag{5.22}$$

where

$$A = \begin{bmatrix} 0_N & I_N \\ -\mathcal{L} & -\gamma\mathcal{L} \end{bmatrix}, \tag{5.23}$$

and $\mathcal{L}$ is the weighted Laplacian matrix with the weight $w_{ij} > 0$, $\gamma > 0$, $b_f^{k\top} = \begin{bmatrix} \bar{b}_f^{k\top} & 0_{1\times N} \end{bmatrix}$ with $\bar{b}_f^k$ being an N by 1 vector with k-th entry as its only nonzero entry. We further assume that node i measures

$$y_i(t) = C_i x(t), \tag{5.24}$$

$C_i = \begin{bmatrix} \bar{C}_i & 0_{|\tilde{\mathcal{N}_i}|\times N} \\ 0_{|\tilde{\mathcal{N}_i}|\times N} & \bar{C}_i \end{bmatrix}$, where $\bar{C}_i$ is a $|\tilde{\mathcal{N}_i}|$ by N matrix of the same structure as considered before. Now we have the following theorem.

Theorem 5.4.4. *Consider the distributed control system with a fault in node k given by (5.22) and local measurments (5.24), and the cases where*

1. $b_f^{k\top} = \begin{bmatrix} \bar{b}_f^{k\top} & 0_{1\times N} \end{bmatrix}$, *or*
2. $b_f^{k\top} = \begin{bmatrix} 0_{1\times N} & \bar{b}_f^{k\top} \end{bmatrix}$

with $\bar{b}_f^k$ being an N by 1 vector with k-th entry as its only nonzero entry. If $\mathcal{G}$ is connected and $k \in N_i$, then there exists a UIO for node i.

Remark 5.4.2. *Proofs of Theorems 5.4.3 and 5.4.4 are similar to the proof of Theorem 5.4.1 and are therefore omitted.*

So far we have established what type of measurements should be available at node i to be able to detect a fault in $k \in \mathcal{N}_i$ using a UIO fault detection scheme. More specifically we have shown that if a node aims to detect a fault in a state of one of its neighbors using a UIO based scheme, it has to measure the same state of all of its neighbors.

Definition 5.4.5 (Monitoring node)**.** *Any node using UIOs to detect and isolate faults in the system is denoted as a* monitoring node*.*

Using the above definition, a first approach to detect and isolate faults in the network is to have each node monitoring its neighbors, i.e., every node is a monitoring node. Although this scheme has high redundancy, it also leads to a highly complex and computational intensive implementation. This issue is further addressed in Section 5.5, where we consider alternative schemes with lower complexity.

In the next section we address the problem of reconfiguring the distributed control law after detecting a fault in the network.

5.4.3 Faulty Node Removal

In this section, we make the following assumptions for the considered graph. We assume that the graph $\mathcal{G}$ is 2-vertex-connected, i.e., after losing any single vertex it remains connected. This results in the graph $\mathcal{G}$ to be also 2-edge-connected, i.e., after losing any single edge it remains connected. Moreover, we consider the case where there is at most one faulty node, k, in the formation and the fault is either in $\xi_k(t)$ or in $\zeta_k(t)$. We propose the algorithm described in Figure 5.1 to solve the problem of automatically reconfiguring the distributed control law to cope with a faulty node.

Now, consider the network described in Section 5.2 with constant external inputs v, where $v \neq 0 \in \mathbb{R}^{2N}$, and the assumptions previously made. Consider the stability of this system where $\dot{x}(t) = Ax(t) + v$. A condition on v for the system to converge to an equilibrium point can be identified (entries of v adds to zero.). Note that the algorithm depicted in Figure 5.1 cannot be applied to remove the faulty node for such a system with a non-zero input. The reason is that if one applies the algorithm after locating the faulty node, v loses one element and the entries of v do not add up to zero anymore, which will drive the system to instability. To remedy this issue, we modify the aforementioned algorithm to deal with removal of the faulty node in such systems, and replace v_ℓ ($\ell \in \mathcal{N}_k$) by $v_\ell + \dfrac{v_k}{|\mathcal{N}_k|}$ after removing the faulty node k to ensure convergence to an equilibrium. This procedure could be seen as assigning the generator nodes in $\mathcal{N}_k$ as distributed slack buses.

5.5 Reducing the Complexity of Distributed FDI Method

For implementation of the distributed FDI method introduced in this chapter, at each node it is required to have one observer corresponding to each of the neighbors. Each of these observers have $2N$ states. So at each node i, $2N|\mathcal{N}_i|$ states are estimated, which puts a heavy computational burden on each of the nodes as N increases. For example a network with 10 nodes the observer bank in node a node with 5 neighbors would require a total of 100 states. So it is desired to reduce the amount of computation necessary for the FDI scheme, which can be achieved in two ways. The first one, is to find a way to decrease the number of monitoring nodes in the network while guaranteeing that each node in the network is being monitored

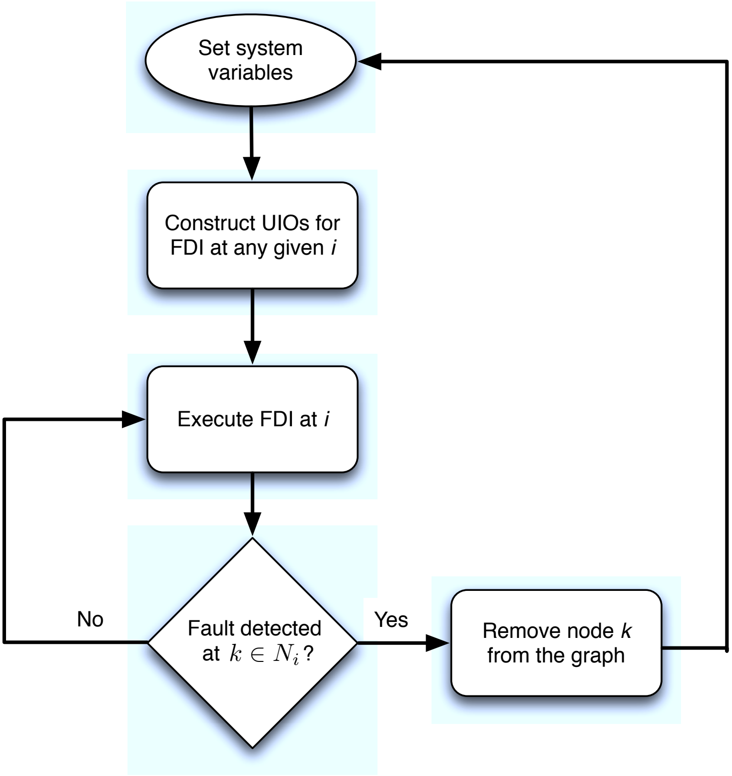

Set system variables
Construct UIOs for FDI at any given i
Execute FDI at i
Fault detected at $k \in N_i$?
No
Yes
Remove node k from the graph

Figure 5.1: Faulty node removal and distributed control law in the presence of fault

by at least another node and calculating UIOs for only these nodes. The second way involves reducing the dimensions of the UIOs used in the distributed FDI.

5.5.1 Reducing the Number of Monitoring Nodes

Considering initially that each monitoring node monitors all the respective neighbors, we say that an FDI system in node i covers the set of nodes $\mathcal{N}_i$. The objective is to select a minimum number of monitoring nodes so that they cover all the nodes in the network, *i.e.*,

$$\begin{array}{ll} \min\limits_{S_o \subseteq \mathcal{V}} & |S_o| \\ \text{s.t.} & \bigcup\limits_{i \in S_o} \mathcal{N}_i = \mathcal{V} \end{array} , \qquad (5.25)$$

where S_o is the set of monitoring nodes. This is actually a set cover problem where we wish to determine a *minimum total dominating set*, *i.e.*, a set with minimum cardinality such that all nodes in the graph have at least one neighbor in that set. This is a well studied problem, having been classified as an NP-hard problem and two algorithms to solve this problem can be found in (Grandoni, 2006).

Although the number of observers obtained by using $\mathcal{N}_i$ as the set of nodes covered by node i is not minimum, this method has one interesting property: all nodes in S_o are monitored by at least one neighbor. This means that even if an observer node is attacked, there is another observer node in the network that can detect it. Obviously, this decreases the vulnerability of such scheme to faults in the monitoring nodes.

Other interesting properties may also be imposed by modifying the constraints in (5.25), such as having S_o to be connected, which is related to the *minimum connected dominating set* problem.

Another way of minimizing the computational burden of the proposed method is to find a set of nodes that monitors all the nodes in the network with the minimum number of measurements, that is solving the following problem:

$$\begin{array}{ll} \min\limits_{S_o \subseteq \mathcal{V}} & \sum\limits_{i \in S_o} \deg(i) \\ \text{s.t.} & \bigcup\limits_{i \in S_o} \mathcal{N}_i = \mathcal{V}. \end{array} , \tag{5.26}$$

This problem can be solved first by finding all the dominating sets in the network and choosing the set that minimizes the cost function.

5.5.2 Reducing the Dimensions of the UIOs

Another interesting approach to reduce the complexity of the distributed FDI method is to use only local models of the system, thus reducing the dimension on each UIO. The main concept of this approach is to design the bank of UIOs at each node i so that the dynamics of each UIO become decoupled from part of the networked system, while ensuring that any fault in the set $\mathcal{N}_i$ can still be detected and isolated. Although this method requires additional measurements from outside the neighbor set, it can achieve a large reduction of the dimension of each UIO. An example illustrating the proposed scheme is presented in the next section and the details for implementing this approach can be found in (Shames et al., 2012).

5.6 Application to Practical Examples

In this section we consider the problem of fault detection and isolation in two practical examples. First we consider detection and isolation of faults in power networks and then we consider the same problem in a formation of mobile nodes with double integrator dynamics.

5.6.1 Distributed FDI in Power Networks

In what follows we propose a fault detection and isolation scheme for a power system akin to the one presented earlier. We assume that all the buses in the network are connected to synchronous machines (motors or generators). The behavior of a synchronous electrical motor located in bus i can be described by the so-called swing equation:

$$m_i\ddot{\delta}_i(t) + d_i\dot{\delta}_i(t) - P_{mi}(t) = -\sum_{j\in\mathcal{N}_i} P_{ij}(t), \tag{5.27}$$

where δ_i is the phase angle of bus i, m_i and d_i are the inertia and damping coefficients, respectively, P_{mi} is the mechanical input power and P_{ij} is the active power flow from bus i to j, see (Guedes et al., 2005). Considering that there are no power losses nor ground admittances and letting $V_i = |V_i|\, e^{j\delta_i}$ be the complex voltage of bus i, the active power flow between bus i and bus j, P_{ij}, is given by:

$$P_{ij}(t) = k_{ij}\sin(\delta_i(t) - \delta_j(t)) \tag{5.28}$$

where $k_{ij} = |V_i|\,|V_j|\, b_{ij}$ and b_{ij} is the susceptance of the power line connecting buses i and j.

Since the phase angles are close, we can linearize (5.28), rewriting the dynamics of bus i as:

$$m_i\ddot{\delta}_i(t) + d_i\dot{\delta}_i(t) = -\sum_{j\in\mathcal{N}_i} k_{ij}(\delta_i(t) - \delta_j(t)) + P_{mi}. \tag{5.29}$$

Consider a power network with $\mathcal{G}(\mathcal{V},\mathcal{E})$ as its underlying graph with $N = |\mathcal{V}|$ nodes, where each node corresponds to a bus in the power network. Rewriting (5.29) in state-state form and considering $x = \left[\delta_1(t),\ \cdots,\ \delta_N(t), \dot{\delta}_1(t),\ \cdots,\ \dot{\delta}_N(t)\right]^\top$ and $v(t) = [P_{m1}\ \cdots\ P_{mN}]^\top$, we have

$$\dot{x}(t) = Ax(t) + Bv(t), \tag{5.30}$$

where $A = \begin{bmatrix} 0_N & I_N \\ -\bar{M}\mathcal{L} & -\bar{M}\bar{D} \end{bmatrix}$, $B = \left[0_N\ \bar{M}\right]^\top$, $\bar{M} = \operatorname{diag}\left(\frac{1}{m_1},\cdots,\frac{1}{m_N}\right)$, $\bar{D} = \operatorname{diag}(d_1,\cdots,d_N)$.

Consider that the network is being affected by faults in the nodes, for instance the removal of a group of generators or loads by local breakers. These result in significant changes on the power generation and consumption and may, under certain conditions, propagate through the network and lead to cascading failures. Assume that a fault has occurred at node k. The power network under such conditions can be modeled as

$$\dot{x}(t) = Ax(t) + Bv(t) + b_f^k f_k, \tag{5.31}$$

where b_f^k is the k-th column of B and therefore it can be written as $b_f^k = \left[0_{1\times N}\ \bar{b}_f^{k\top}\right]^\top$ with $\bar{b}_f^{k\top}$ being a column vector with $\frac{1}{m_k}$ in the k-th entry and zero in all other

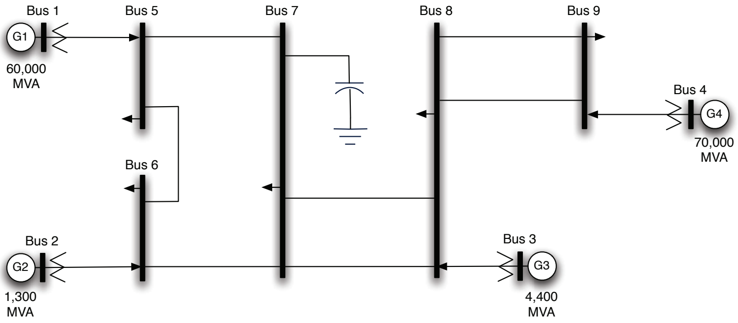


Figure 5.2: Power network with 9 buses (Anderson and Farmer, 1996).

entries. Thus, from Theorem 5.4.3 there exists a UIO for such system at a given node i if $k \in \mathcal{N}_i$ and $y_i = C_i x$ with

$$C_i = \begin{bmatrix} \bar{C}_i & 0_{|\tilde{\mathcal{N}}_i| \times N} \\ 0_{|\tilde{\mathcal{N}}_i| \times N} & \bar{C}_i \end{bmatrix}. \tag{5.32}$$

Thus we need to measure the phase and frequency of the neighbors to be able to detect the faulty node. These measurements are readily available through phase measurement units (PMU). Having such measurements, this type of faults can be detected and isolated in a distributed way using UIOs.

Remark 5.6.1. *Because of Theorem 5.4.2 we know that we cannot solve the fault detection problem using UIO with having access to less information than the information available through $y_i = C_i x$, with the above-mentioned C_i.*

Remark 5.6.2. *In the case where there are buses that are not connected to synchronous machines and are described by algebraic equations; one has two alternatives. First, one can use equation (5.27) to model only the buses that are connected to synchronous machines and use the techniques in (Machowski et al., 2008), Chapter 14, to remove the algebraic relations from the power network model and assume that the faults only affect the buses connected to synchronous machines. Second, one may assume that the buses that are not connected to the machines are governed by dynamic equations of type (5.27), albeit with small damping and inertia coefficients (Guedes et al., 2005).*

Consider the power network presented in Figure 5.2 with four generators in buses 1 to 4 connected through the transmission network to load buses, 5 to 9. The power grid's topological parameters and the generators' dynamic coefficients (m_i and d_i) were taken from (Anderson and Farmer, 1996), while the dynamic coefficients of the rest of the buses were arbitrarily taken from reasonable values.

The power network is evolving towards the steady-state when, at time instant $t = 2s$, a fault occurs at node 6, as presented in Figure 5.3(a). A bank of observers is implemented at bus 7 and the respective residuals are illustrated in Figure 5.3(b). Since the residual corresponding to bus 6, $r_7^6(t)$, remained smaller than the other residuals, the fault is successfully detected and isolated even in the presence of measurement noise.

Distributed FDI using local models

In this section we illustrate the solution proposed in Section 5.5.2 with a power network example. The simulations were carried out using the IEEE 118 bus network available with the MATPOWER toolbox (Zimmerman et al., 2009). The one-line diagram of the power network is depicted in Figure 5.4 and the respective graph is shown in Figure 5.5.

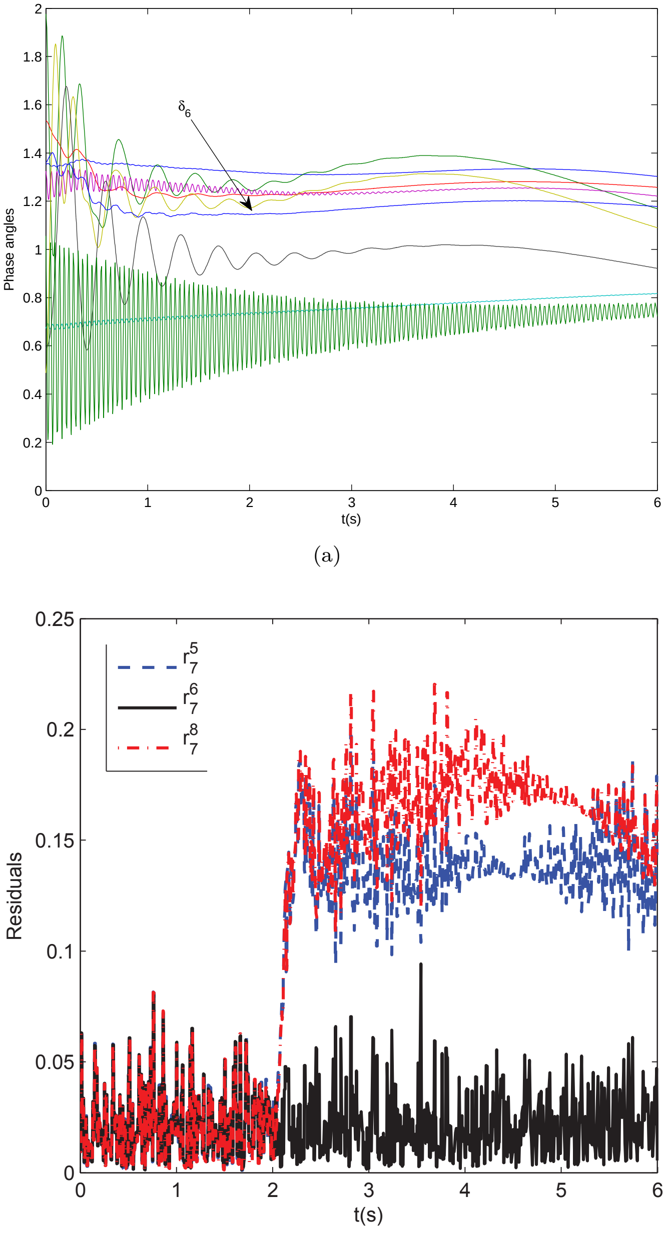
We considered the classical synchronous machine model (Kundur, 1994) for each node of the power network, leading to the global network dynamics as in (5.31). Since the inertia and damping coefficients, m_i and d_i, were not available in the example's data files, they were randomly generated so that the load buses had considerably lower values than the generator buses, namely $m_g \approx 10^3 m_l$ and $d_g \approx 10^3 d_l$.

In this example, node 19 is monitoring its $1-$hop neighbors for faulty behaviors using the method proposed in Section 5.5.2. Thus the only network model knowledge needed is its $2-$hop neighborhood, the smaller cluster in Figure 5.5, which consists of 26 states, as opposed to the 236 states of the global network. Using this smaller model, a bank of UIO was generated according to the discussion in Section 5.5.2.

In the simulations, node 15 exhibits a faulty behavior after $t = 20s$, which is successfully detected by node 19 as seen in Figure 5.6. Furthermore, all the residuals corresponding to other neighboring nodes become large while the one for node 15 remains at zero. Following algorithm in Figure 5.1, node 15 is then detected and identified as the faulty node.

5.6.2 Distributed FDI in Formations of Mobile Agents

In this section adopt the system and the notations introduced in Section 5.4.2. Furthermore, assume at time t_f a fault occurs at node k, one can detect and isolate this fault using the methods introduced earlier. Consider a formation consisting of 10 nodes with double integrator dynamics with the aforementioned control law as depicted in Figure 5.7(a). Further assume at time $t_f = 2$ node 3 starts to malfunction. Using UIOs and the logic presented in (5.13) this fault is detected at time $t_d = 3.56$. A sample of residuals as calculated in node 1 (neighboring node 3) is presented in Figure 5.7(b). In the case where no isolation is carried out the first coordinate of the velocities of the nodes are presented in Figure 5.7(c). However, if after the detection of the fault, the aforementioned algorithm is used



(a)

(b)

Figure 5.3: FDI in a Power Network: (a) Phase angles of the power network. (b) Residuals of buses neighboring bus 7.

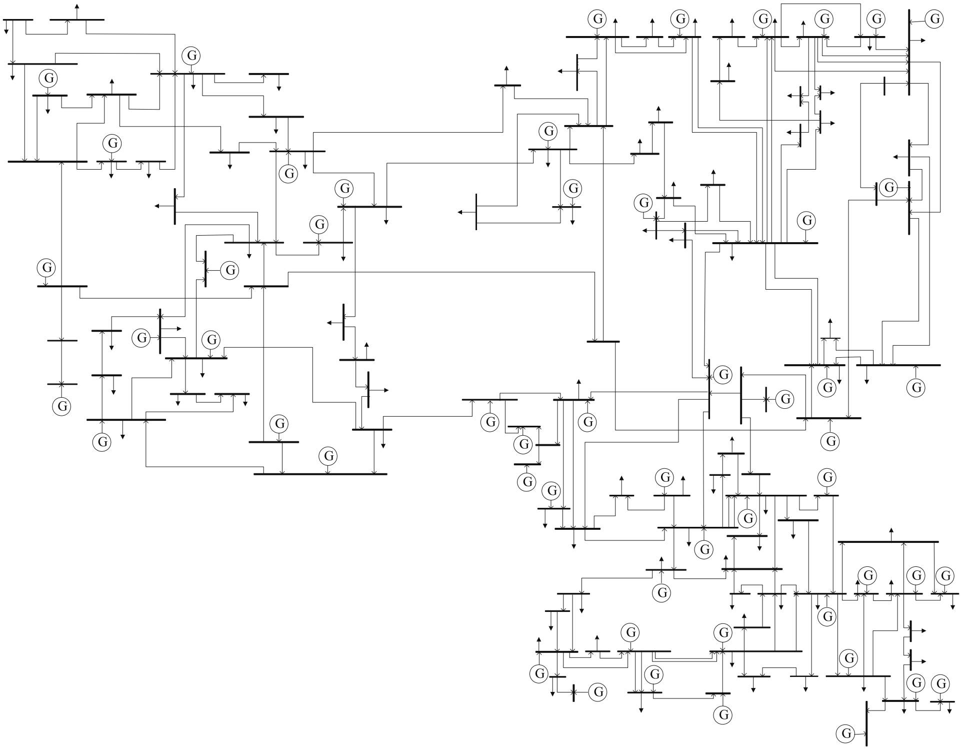


Figure 5.4: One-line diagram of the IEEE 118 bus network (courtesy of the IIT Power Group).

to remove the faulty node, the first coordinate velocities of the nodes would be as the ones depicted in Figure 5.7(d), showing that they have reached consensus. Due to absence of any external input it is not needed to adjust external input after disconnection.

5.7 Summary

In this chapter we considered the problem of fault detection and isolation in the networks of interconnected nodes with double integrator dynamics. A distributed FDI scheme based on UIOs was proposed, requiring only local measurements. Furthermore we analyzed the feasibility of such scheme with respect to local measurements and we also provided some infeasibility results. As part of a mitigation procedure, we proposed an algorithm to remove the faulty node from the network that can also be applied when there are nonzero external inputs. Then we presented some simulation examples related to the motivating applications, thus demonstrating

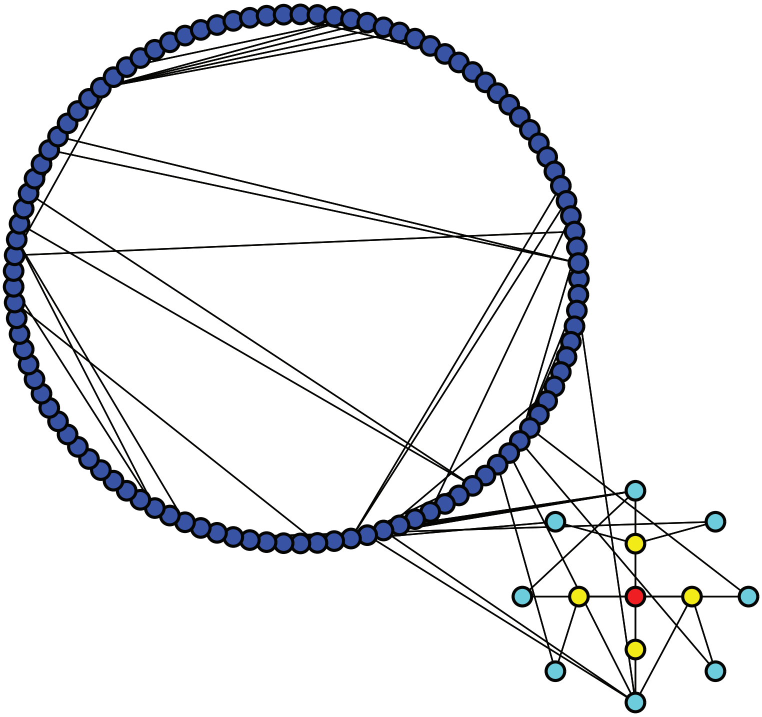


Figure 5.5: The graph of the IEEE 118 bus network. Node 19 is the node at the center of the smaller cluster to the right, monitoring its $1-$hop neighbors. This cluster represents the $2-$hop neighborhood of node 19, where the outer nodes are the $2-$hop neighbors.

the application of the proposed method to fault detection in power and multi-node systems. Some considerations on the complexity and scalability of the proposed method were also given, and possible solutions were indicated.

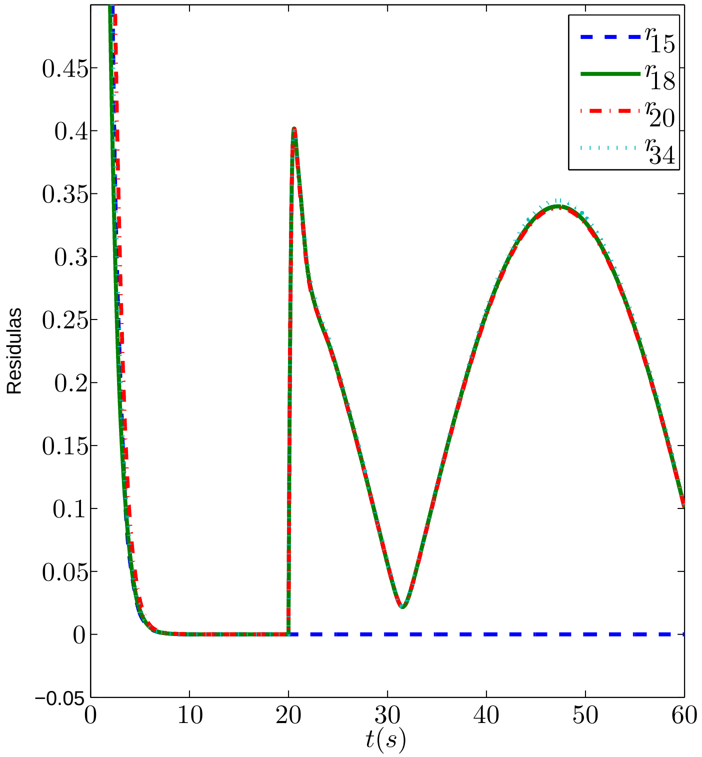
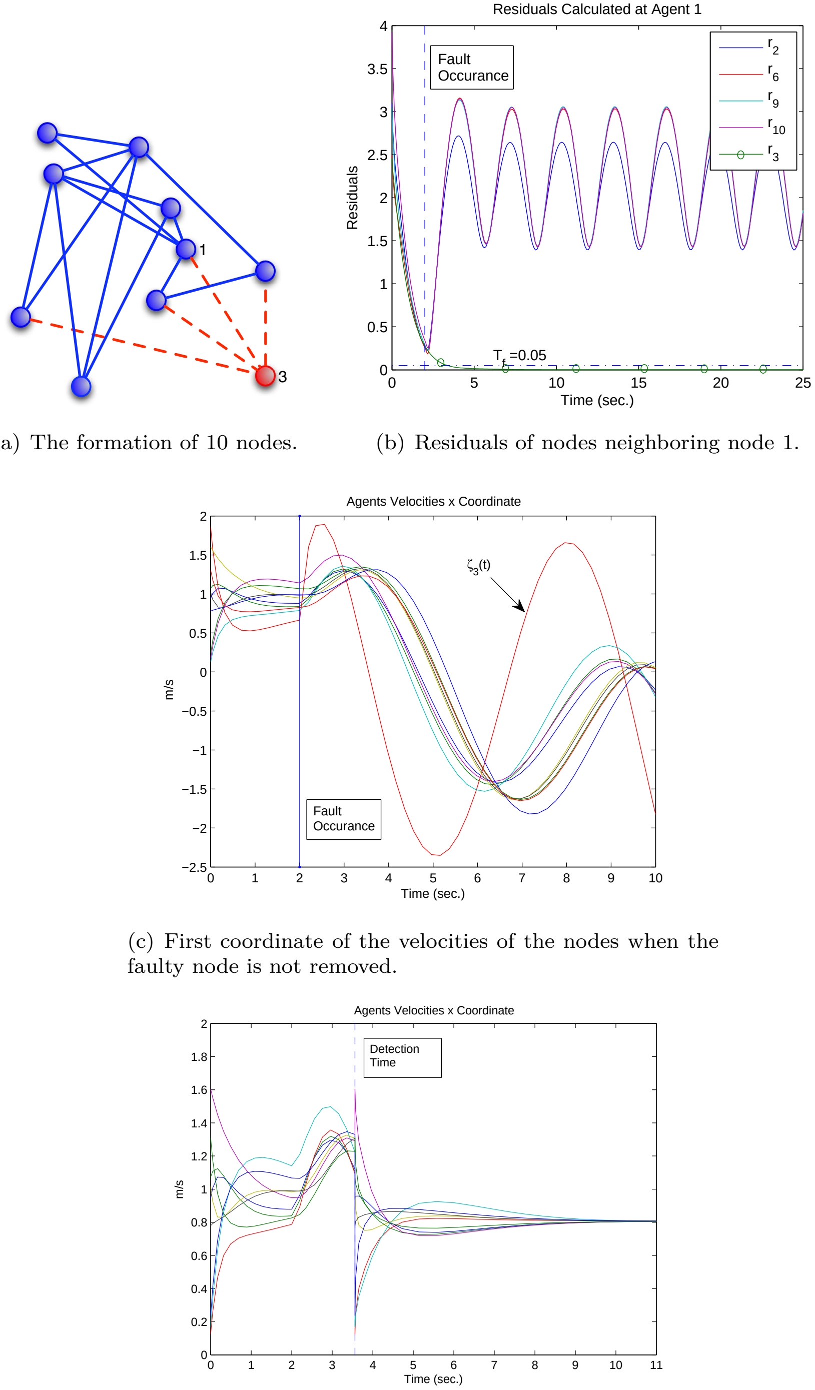


Figure 5.6: Residuals generated by the UIO bank at node 19.



(a) The formation of 10 nodes.

(b) Residuals of nodes neighboring node 1.

(c) First coordinate of the velocities of the nodes when the faulty node is not removed.

(d) First coordinate of the velocities of the nodes when the faulty node is removed.

Figure 5.7: Multi-agent formation in the presence of a fault occurring at $t = 2sec.$ in node 3.

Chapter 6

Conclusions and Future Work

This thesis considered the security and reliability of NCS. Regarding security, this thesis analyzed the power transmission network operation under deception attacks on the measurement data. Stealthy deception attacks and their effects on the OPF and power network operation were characterized. A novel impact-aware security metric was also proposed. Reliability of power transmission networks was also considered by addressing the problem of distributed FDI for interconnected second-order systems. Feasibility of the proposed method was shown and methods to further reduce the complexity of the solution were discussed.

A brief summary of the thesis contributions and possible future research directions are discussed below.

6.1 Conclusions

In this thesis the cyber security of SCADA EMS systems operating power transmission networks under data deception attacks was studied. Considering the attacker to be rational in a special sense, a novel attacker model was proposed, including relevant factors such as attack goals, resource constraints, and a given knowledge of the linearized system model. Former results characterized stealthy attacks for simplified linear models and linear SE algorithms. In this thesis we extend the results to the nonlinear case. The set of stealthy attacks following policies with perfect and perturbed linear model knowledge were characterized using the geometric properties of the linearized models and the detection schemes. Furthermore, the validity of the proposed framework was verified through practical experiments on nonlinear SE algorithms. In spite of the power network being nonlinear, the results obtained by this experiment show that computations based on linear models provide valid stealthy attacks. These attacks successfully corrupt the target measurements without triggering any BDD alarms.

Given the classes of stealthy attacks, their effect on the optimal power flow algorithm and power network operation was studied. Analytical expressions for the fictitious and true profit given the attack were provided. These were used together with the attacker model to propose a novel impact-aware security measurement for each measurement. This security metric considers both the impact on the system and the attack effort, i.e., the number of measurements that are compromised. Therefore, it can be used to develop improved protective schemes.

The power transmission network reliability was also addressed. In particular, we considered the problem of distributed fault detection and isolation in networks of interconnected nodes with double integrator dynamics, corresponding to simplified models of power networks. A distributed FDI scheme based on UIOs was proposed and the feasibility of the approach was analyzed with respect to local measurements. Some infeasibility results were also provided. The complexity of the proposed scheme was also analyzed, showing that it may not be scalable. Methods to reduce the complexity of the proposed scheme were discussed, thus ensuring the scalability of the solution.

6.2 Future Work

There are several research directions to explore regarding the work presented in this thesis and security and reliability of NCS in general. In this section we discuss some of these directions for possible future work.

Chapter 3: Cyber Security of State Estimator in Power Systems

This thesis characterized stealthy deception attacks for SEs using the Gauss-Newton method and standard BDD schemes with both perfect and perturbed linear models. However, this study was based on the assumption that the SE converged. Although the practical experiments showed that stealthy attacks based on the DC network model can significantly corrupt the estimates while having the SE to converge, it remains unclear what the constraints on the attack magnitude and the perturbed model are for which convergence is ensured.

Alternative SE methods suggested in the literature could also be analyzed under deception attacks. For instance, there are the so-called robust SE which are known to be more robust to outliers than the WLS approach, so studying them regarding attacks might give further insights on the SE security.

Chapter 4: Cyber Security of Optimal Power Flow in Power Systems

The proposed impact-aware security metric is posed as a non-convex optimization problem. Therefore heuristics for its efficient computation are desired. Moreover, it was previously mentioned that the novel impact-aware security metric could be used for protective schemes. Designing these schemes is of interest to mitigate the effects of deceptions attacks on power network operation.

Chapter 5: Distributed Fault Diagnosis in Networked Systems

The proposed distributed FDI scheme was applied to power transmission networks modeled by the swing-equation. However, more detailed power system models exist and distributed FDI for these system still remains an open problem. Additionally, similarly to the approach in Chapter 3, the distributed FDI scheme could be a target of cyber attacks. However, attacks on dynamical system are much harder than on static systems, since time and the dynamic models provide additional information. It is therefore relevant to analyze these schemes under possible cyber attacks scenarios.

Secure and Reliable NCS

In Section 2.3 an overview of recent research on secure and reliable NCS was presented. Several problems in this area remain unsolved, as also seen in the previous sections. Moreover, sound frameworks to study security and reliability of NCS are still needed.

Bibliography

A. Abur and A.G. Exposito. *Power System State Estimation: Theory and Implementation.* Marcel-Dekker, 2004.

M. Aldeen and F. Crusca. Observer-based fault detection and identification scheme for power systems. *IEE Proceedings - Generation, Transmission and Distribution*, 153(1):71–79, January 2006.

S. Amin, A.A. Cárdenas, and S.S. Sastry. Safe and secure networked control systems under denial-of-service attacks. In *Hybrid Systems: Computation and Control*, pages 31–45. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, April 2009.

S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen. Stealthy deception attacks on water scada systems. In *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control*, HSCC '10, pages 161–170, New York, NY, USA, 2010. ACM.

P. M. Anderson and R. G. Farmer. *Series compensation of power systems.* PBLSH. Inc, California, USA, 1996.

J. Baillieul and P. J. Antsaklis. Control and Communication Challenges in Networked Real-Time Systems. *Proceedings of the IEEE*, 95(1):9–28, January 2007. ISSN 0018-9219. doi: 10.1109/JPROC.2006.887290.

N. Balu, T. Bertram, A. Bose, V. Brandwajn, G. Cauley, D. Curtice, A. Fouad, L. Fink, M.G. Lauby, B.F. Wollenberg, and J.N. Wrubel. On-line power system security analysis. *Proceedings of the IEEE*, 80(2):262–282, feb 1992.

P. Barooah and J.P. Hespanha. Graph effective resistance and distributed control: Spectral properties and applications. In *Decision and Control, 2006 45th IEEE Conference on*, pages 3479–3485. IEEE, 2007.

M. Basseville and I. V. Nikiforov. *Detection of abrupt changes: theory and application.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

M. Bishop. *Computer Security: Art and Science.* Addison-Wesley Professional, 2002.

G. Björkman. The VIKING project–torwards more secure SCADA systems. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010*, Stockholm, Sweden, April 2010.

R. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. Overbye. Detecting false data injection attacks on DC state estimation. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010*, Stockholm, Sweden, April 2010.

S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry. Challenges for securing cyber physical systems. In *Workshop on Future Directions in Cyber-physical Systems Security.* DHS, July 2009.

A. Cárdenas, S. Amin, Z. Lin, Y. Huang, C. Huang, and S. Sastry. Attacks against process control systems: risk assessment, detection, and response. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '11, pages 355–366, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0564-8.

A. A. Cárdenas, S. Amin, and S. S. Sastry. Secure control: Towards survivable cyber-physical systems. In *First International Workshop on Cyber-Physical Systems (WCPS2008)*, pages 495–500, Beijing, China, June 2008a. URL http://www.truststc.org/pubs/345.html.

A.A. Cárdenas, S. Amin, and S.S. Sastry. Research challenges for the security of control systems. In *Proc. 3rd USENIX Workshop on Hot topics in security*, July 2008b.

CBSNews. Cyber war: Sabotaging the system. *CBSNews*, November 8th 2009.

J. Chen and R. J. Patton. *Robust Model-Based Fault Diagnosis for Dynamic Systems.* Kluwer Academic Publishers, 1999.

J. Chen, R. J. Patton, and H. Zhang. Design of unknown input observers and robust fault detection filters. *International Journal of Control*, 63(1):85–105, 1996.

Y. Chompoobutrgool, L. Vanfretti, and M. Ghandhari. Survey on power system stabilizers control and their prospective applications for power system damping using synchrophasor-based wide-area systems. *European Transactions on Electrical Power*, 2011.

E. Chow and A. Willsky. Analytical redundancy and the design of robust failure detection systems. *Automatic Control, IEEE Transactions on*, 29(7):603–614, jul 1984.

W. H. Chung, J. L. Speyer, and R. H. Chen. A decentralized fault detection filter. *Journal of Dynamic Systems, Measurement, and Control*, 123(2):237–247, 2001.

W.H. Chung and J.L. Speyer. A game theoretic fault detection filter. *Automatic Control, IEEE Transactions on*, 43(2):143–161, feb 1998.

G. Dán and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *Proc. of IEEE SmartGridComm*, October 2010.

M. A. Demetriou. Using unknown input observers for robust adaptive fault detection in vector second-order systems. *Mechanical systems and signal processing*, 19(2):291–309, 2005. ISSN 0888-3270.

S. X. Ding. *Model-based Fault Diagnosis Techniques: Design Schemes.* Springer Verlag, 2008.

S. X. Ding, P. Zhang, Ch. Chihaia, W. Li, Y. Wang, and E. L. Ding. Advanced design scheme for fault tolerant distributed networked control systems. In *Proceedings of the 17th IFAC World Congress*, pages 13569 – 13574, Seoul, Korea, July 2008.

R.K. Douglas and J.l. Speyer. Robust fault detection filter design. In *American Control Conference, 1995. Proceedings of the*, volume 1, pages 91–96, jun 1995.

P. 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, 2010*, pages 962–967, jul 2010.

N. Falliere, L. Murchu, and E. Chien. W32.Stuxnet dossier, February 2011.

FERC. Final report on price manipulation in western markets, March 2003. Available at: http://www.ferc.gov/industries/electric/indus-act/wec.asp.

P. M. Frank and X. Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of process control*, 7 (6):403–424, 1997.

A. Galántai. Subspaces, angles and pairs of orthogonal projections. *Linear and Multilinear Algebra*, 56(3):227–260, June 2006.

A. Giani, S. Sastry, K. H. Johansson, and H. Sandberg. The VIKING project: an initiative on resilient control of power networks. In *Proc. 2nd Int. Symp. on Resilient Control Systems*, pages 31–35, Idaho Falls, ID, USA, August 2009.

A. Giani, E. Bitar, M. McQueen, P. Khargonekar, K. Poolla, and M. Garcia. Smart grid data integrity attacks: Characterizations and countermeasures. In *Proceedings of the IEEE SmartGridComm*, October 2011. To appear.

S. Gorman. Electricity grid in U.S. penetrated by spies. *The Wall Street Journal*, page A1, April 8th 2009.

F. Grandoni. A note on the complexity of minimum dominating set. *J. Discrete Algorithms*, 4(2):209–214, July 2006. URL http://dblp.uni-trier.de/db/journals/jda/jda4.html#Grandoni06.

R. B. L. Guedes, F. Silva, L. F. C. Alberto, and N. G. Bretas. Large disturbance voltage stability assessment using extended Lyapunov function and considering voltage dependent active loads. In *Power Engineering Society General Meeting, 2005. IEEE*, pages 1760–1767. IEEE, 2005.

A. Gupta, C. Langbort, and T. Başar. Optimal control in the presence of an intelligent jammer with limited actions. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 1096–1101, dec 2010.

J.P. Hespanha, P. Naghshtabrizi, and Yonggang Xu. A survey of recent results in networked control systems. *Proceedings of the IEEE*, 95(1):138–162, jan 2007.

I. Hwang, S. Kim, Y. Kim, and C. E. Seah. A survey of fault detection, isolation, and reconfiguration methods. *Control Systems Technology, IEEE Transactions on*, 18(3):636–653, may 2010.

R. Isermann. Model-based fault detection and diagnosis: status and applications. In *Proceedings of the 16th IFAC Symposium on Automatic Control in Aerospace*, pages 71–85, St. Petersburg, Russia, June 2004.

L. Jia, R. J. Thomas, and L. Tong. Malicious data attack on real-time electricity market. In *Proc. of IEEE ICASSP*, May 2011.

O. Kosut, L. Jia, R. Thomas, and L. Tong. Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. In *Proc. of IEEE SmartGridComm*, October 2010.

P. Kundur. *Power System Stability and Control*. McGraw-Hill Professional, 1994.

Wen-Hsiung E. Liu, Felix F. Wu, and Shau-Ming Lun. Estimation of parameter errors from measurement residuals in state estimation. *IEEE Transactions on Power Systems*, (1), February 1992.

Y. Liu, M. K. Reiter, and P. Ning. False data injection attacks against state estimation in electric power grids. In *Proc. 16th ACM Conf. on Computer and Communications Security*, pages 21–32, New York, NY, USA, 2009.

J. Machowski, J. W. Bialek, and J. R. Bumby. *Power System Dynamics: Stability and Control*. John Wiley & Sons, 2008.

J. Meserve. Sources: Staged cyber attack reveals vulnerability in power grid. *CNN*, 2007. Available at http://edition.cnn.com/2007/US/09/26/power.at.risk/index.html.

L. Mili, Th. Van Cutsem, and M. Ribbens-Pavella. Bad data identification methods in power system state estimation - a comparative study. In *IEEE Trans. Power App. Syst.*, November 1985.

Y. Mo and B. Sinopoli. Secure control against replay attack. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 911–918, October 2009.

A. Monticelli. *State Estimation in Electric Power Systems: A Generalized Approach*. Kluwer Academic Publishers, 1999.

R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, 1982.

F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, 2010. Submitted, available online at http://www.fabiopas.it/papers/FP-AB-FB-10a.pdf.

F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. *Automatic Control, IEEE Transactions on*, PP(99):1, 2011.

R. J. Patton and J. Chen. Observer-based fault detection and isolation: robustness and applications. *Control Engineering Practice*, 5(5):671–682, 1997.

A.G. Phadke and R.M. de Moraes. The wide world of wide-area measurement. *Power and Energy Magazine, IEEE*, 6(5):52 –65, sep-oct 2008. ISSN 1540-7977.

T. Rid. Cyber war will not take place. *Journal of Strategic Studies*, 2011. doi: 10.1080/01402390.2011.608939.

C.G. Rieger. Notional examples and benchmark aspects of a resilient control system. In *Resilient Control Systems (ISRCS), 3rd International Symposium on*, pages 64–71, aug 2010. doi: 10.1109/ISRCS.2010.5603123.

T. Samad and A.M. Annaswamy, editors. *The Impact of Control Technology*. IEEE Control Systems Society, 2011. Available at http://www.ieeecss.org/general/impact-control-technology.

T. Samad, P. McLaughlin, and J. Lu. System architecture for process automation: Review and trends. *Journal of Process Control*, 17(3):191–201, 2007. Special Issue ADCHEM 2006 Symposium.

H. Sandberg, A. Teixeira, and K. H. Johansson. On security indices for state estimators in power networks. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010*, Stockholm, Sweden, April 2010.

E. Scholtz and B.C. Lesieutre. Graphical observer design suitable for large-scale DAE power systems. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 2955–2960, Cancun, December 2008.

F. C. Schweppe and J. Wildes. Power system static-state estimation, part I: Exact model. *IEEE Transactions on Power Apparatus and Systems*, 89(1):120–125, January 1970.

M. Shahidehpour, F. Tinney, and Y. Fu. Impact of security on power systems operation. *Proceedings of the IEEE*, 93(11):2013–2025, nov 2005a.

M. Shahidehpour, W. F. Tinney, and Y. Fu. Impact of security on power systems operation. *Proceedings of the IEEE*, 93(11):2013–2025, November 2005b.

I. Shames, A. Teixeira, H. Sandberg, and K. H. Johansson. Distributed fault detection and isolation with imprecise network models. In *American Control Conference*, 2012. Submitted.

D. D. Siljak. *Decentralized control of complex systems*. Academic Press, 1991.

R. Smith. A decoupled feedback structure for covertly appropriating networked control systems. In *Proceedings of the 18th IFAC World Congress*, Milano, Italy, August-September 2011.

K. C. Sou, H. Sandberg, and K. H. Johansson. Electric power network security analysis via minimum cut relaxation. In *Proceedings of the 50th IEEE Conference on Decision and Control*, December 2011. To appear.

S. Sundaram and C.N. Hadjicostis. Distributed function calculation via linear iterative strategies in the presence of malicious agents. *Automatic Control, IEEE Transactions on*, 56(7):1495–1508, july 2011.

S. Sundaram, M. Pajic, C.N. Hadjicostis, R. Mangharam, and G.J. Pappas. The wireless control network: Monitoring for malicious behavior. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 5979–5984, dec 2010.

SvK. Map of sweden's national grid.

Symantec. W32.Duqu: The precursor to the next stuxnet, November 2011.

U.S.-Canada PSOTF. Final report on the August 14th blackout in the United States and Canada. Technical report, U.S.-Canada Power System Outage Task Force, April 2004.

F. F. Wu. Power system state estimation: a survey. *Int. J. Elec. Power and Energy Systems*, April 1990.

F. F. Wu and W.E. Liu. Detection of topology errors by state estimation. *IEEE Trans. Power Syst.*, (1), February 1989.

F.F. Wu, K. Moslehi, and A. Bose. Power system control centers: Past, present, and future. *Proceedings of the IEEE*, 93(11):1890–1908, nov 2005. ISSN 0018-9219.

L. Xie, Y. Mo, and B. Sinopoli. False data injection attacks in electricity markets. In *First IEEE International Conference on Smart Grid Communications*, October 2010.

Y. Yuan, Z. Li, and K. Ren. Modeling load redistribution attacks in power systems. *IEEE Transactions on Smart Grid*, 2(2):382–390, June 2011.

W. Zhang, Q. Yang, and Y. Geng. A survey of anomaly detection methods in networks. In *International Symposium on Computer Network and Multimedia Technology*, 2009.

R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas. MATPOWER's extensible optimal power flow architecture. In *Power and Energy Society General Meeting*, pages 1–7. IEEE, July 2009.