DATA-DRIVEN METHODS TO IMPROVE RESOURCE UTILIZATION, FRAUD DETECTION, AND CYBER-RESILIENCE IN SMART GRIDS

BY

VARUN BADRINATH KRISHNA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor William H. Sanders, Chair
Professor Carl A. Gunter
Professor Ravishankar K. Iyer
Professor Klara Nahrstedt
Professor Peter Sauer

ABSTRACT

This dissertation demonstrates that empirical models of generation and consumption, constructed using machine learning and statistical methods, improve resource utilization, fraud detection, and cyber-resilience in smart grids.

The modern power grid, known as the *smart grid*, uses computer communication networks to improve efficiency by transporting control and monitoring messages between devices. At a high level, those messages aid in ensuring that power generation meets the constantly changing power demand in a manner that minimizes costs to the stakeholders. In buildings, or *nanogrids*, communications between loads and centralized controls allow for more efficient electricity use. Ultimately, all efficiency improvements are enabled by data, and it is vital to protect the integrity of the data because compromised data could undermine those improvements. Furthermore, such compromise could have both economic consequences, such as power theft, and safety-critical consequences, such as blackouts.

This dissertation addresses three concerns related to the smart grid: resource utilization, fraud detection, and cyber-resilience. We describe energy resource utilization benefits that can be achieved by using machine learning for renewable energy integration and also for energy management of building loads. In the context of fraud detection, we present a framework for identifying attacks that aim to make fraudulent monetary gains by compromising consumption and generation readings taken by meters. We then present machine learning, signal processing, and information-theoretic approaches for mitigating those attacks. Finally, we explore attacks that seek to undermine the resilience of the grid to faults by compromising generators' ability to compensate for lost generation elsewhere in the grid. Redundant sources of measurements are used to detect such attacks by identifying mismatches between expected and measured behavior.

To my late grandfathers—K. Krishna Ayengar and B. N. Parthasarathy

ACKNOWLEDGMENTS

Several people played vital roles in ensuring that I had an impactful, productive, and enjoyable time while I pursued a Ph.D. at the University of Illinois at Urbana-Champaign. I would first like to thank my Ph.D. adviser Prof. William H. Sanders for giving me the opportunity to work with him in the best smart grid security research program in the world. In addition to giving me research support, he always believed in my ideas and in my ability to execute those ideas. He also trusted me to represent him and the university by giving me many opportunities to present our joint research to audiences from academia and industry. Furthermore, he gave me plenty of career-related guidance and advice. I would also like to thank the following people for specific ways in which they helped me on my journey:

- Prof. Sanders, Prof. Ravishankar Iyer, Prof. Carl A. Gunter, Prof. Peter Sauer, and Prof. Klara Nahrstedt, for their guidance as members of my doctoral committee.

- Prof. Sanders, Prof. Iyer, Prof. Gunter, Prof. Zbigniew Kalbarcyk, Prof. David Yau, Dr. Deokwoo Jung, Prof. Rui Tan, Dr. Ziping Wu, Dr. Kiryung Lee, Dr. Wander Wadman, Dr. Younghun Kim, Michael Rausch, Richard Macwan, Boya Hou, Peng Gu, Vaidehi Ambardekar, Hoang Hai Nguyen, William Temple, and Ng Quo Khiem, with whom I collaborated on the work included in this dissertation.

- Tim Yardley, Dr. Ziping Wu, Richard Macwan, Prosper Panumpabi, and Jeremy Jones for support with testbed resources.

- Prof. Marianne Winslett, Prof. Douglas L. Jones, Prof. Iyer, Prof. Kalbarcyk, and Prof. Yih Chun Hu, for helping me prepare my graduate school application.

- Prof. Sanders, Prof. Iyer, Prof. Winslett, Prof. Yau, and Dr. Bimlesh Wadhwa, for

writing letters of recommendation for graduate school and various fellowships, awards, and internships during graduate school.

- Prof. Ashwin M. Khambadkone for giving me my first opportunity to work on smart grids at the Experimental Power Grid Centre (EGPC) at A*STAR.

- Prof. David Nicol, Prof. Michael Bailey, and Prof. Yih Chun Hu for passing me on my Ph.D. qualifying exam.

- My colleagues in the PERFORM Group: Ahmed Fawaz, Atul Bohara, Ben Ujcich, Brett Feddersen, Carmen Cheh, Ken Keefe, Michael Rausch, Mohammad Noureddine, Ron Wright, and Uttam Thakore for sharing food and food for thought.

- Members and administrators of the Trustworthy Cyber Infrastructure for the Power Grid (TCIPG) and Cyber-Resilient Energy Delivery Consortium (CREDC) centers.

- Other colleagues in the Information Trust Institute, ADSC, and EPGC from whom I have learned a lot.

- Jenny Applequist and James Hutchinson for their help with improving the writing quality of the dissertation.

- My friends at UIUC, Aarti Shah, Linjia Chang, Ishita Bisht, Erik Johnson, Sarah Robinson, Daniel Fisher, Chelsea Fry, James Pikul, Anthony Christodoulou, Kartik Palani, Vignesh Babu, Rakesh Kumar, Sakshi Srivastava, Shweta Patwa, Snegha Ramnarayan, Kato Lindholm, Alex Tecza, Akshat Puri, Yana Garmash, Jeffrey Proulx, Cassie Liu, Nicole Cox, Jonathan Lai, Kirill Mangutov, Anna Kalinowski, John Hadley, Anna Vardanyan, David Meldgin, Laleh Omarie, Aneysha Bhat, Samantha Soukup, Danica Fong, Arjun Athreya, Subho Banerjee, Yoga Varatharajah, Krishnakanth Saboo, Saurabh Jha, Homa Alemzahed, Ashwarya Rajvardhan, Siddhanth Munukutla, Arunita Kar, Pei Han, Mei Ling, and others for ensuring that I had a good social life outside of research during the years of my Ph.D.

- Dr. Eric Davis, Prof. Rakesh Kumar, and Prof. Lav Varshney for career advice.

- The creators of icons used in Figs. 1.2 and 7.2 and obtained under Flaticon Free License: Freepik, Icon Pond, Skyclick, Srip, Smashicons, and Those Icons.

- And last, but not least, my family for providing me with emotional support, cooking advice, career advice, and all kinds of life advice.

Funding Sources

Most of the work in this dissertation was supported by the U.S. Department of Energy (DOE) and Department of Homeland Security (DHS) under the Trustworthy Cyber Infrastructure for the Power Grid (TCIPG) and Cyber-Resilient Energy Delivery Consortium (CREDC) centers. We thank Dr. Carol Hawk for her support through spearheading of these DOE initiatives.

- Chapter 2 was partially supported by the IBM T.J. Watson Research Center, Utopus Insights Inc. and the U.S. Department of Energy under Award Number DE-OE0000780 (CREDC).

- Chapter 3 was partially supported by Singapore's Agency for Science, Technology, and Research (A∗STAR), through a research grant for the Human Sixth Sense Programme at the Advanced Digital Sciences Center. The sensor-network work for quantifying energy use and demand response was partially funded by the Korea Micro Energy Grid (KMEG) of the Office of Strategic R&D Planning (OSP), Korea government Ministry of Knowledge Economy (No. 2011T100100024). The evaluation of attacks on real-time pricing was partially funded by the U.S. National Science Foundation under grant numbers CNS-0963715 and CNS-0964086.

- Chapters 4, 5, and 6 were partially supported by the Siebel Energy Institute and the U.S. Department of Energy under Award Numbers DE-OE0000097 (TCIPG).

- Chapter 7 was partially supported by the U.S. Department of Energy under Award Number DE-OE0000780 (CREDC).

Disclaimer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

2-D	Two-dimensional
3-D	Three-dimensional
ACF	Autocorrelation Function
ADSC	Advanced Digital Sciences Center
AGC	Automatic Generation Control
AMI	Advanced Metering Infrastructure
APT	Advanced Persistent Threat
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARX	Autoregressive model with eXogenous inputs
AUC	Area Under the Curve
BC	British Columbia
CDF	Cumulative Density Function
CER	Ireland's Commission for Energy Regulation
CIP	Critical Infrastructure Protection
CPU	Central Processing Unit
CREDC	Cyber-Resilient Energy Delivery Consortium
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DCS	Distributed Control Systems

DER	Distributed Energy Resource
DHS	U.S. Department of Homeland Security
DOE	U.S. Department of Energy
ECE	Electrical and Computer Engineering
FP	False Positive
FPR	False Positive Rate
GE	General Electric Company
HVAC	Heating, Ventilation, and Air Conditioning
IEEE	Institute of Electrical and Electronics Engineers
IT	Information Technology
ITI	Information Trust Institute
KLD	Kullback-Leibler Divergence
MAE	Mean Absolute Error
NARX	Non-linear Auto-Regressive model with eXogenous inputs
NERC	North American Electric Reliability Corporation
NREL	National Renewable Energy Laboratory
OT	Operational Technology
PCA	Principal Component Analysis
PCNN	Power Curve Neural Network
PDF	Probability Density Function
PFR	Primary Frequency Response
REG	Renewable Energy Generator
ROC	Receiver Operating Characteristic
RTP	Real-time Pricing
SFR	Secondary Frequency Response
TCIPG	Trustworthy Cyber Infrastructure for the Power Grid
TOU	Time-of-Use

TP	True Positive
TPR	True Positive Rate
UFLS	Under-Frequency Load Shedding
UIUC	University of Illinois at Urbana-Champaign
VELCO	Vermont Electric Power Company
WECC	Western Electric Coordinating Council
WTG	Wind Turbine Generator
ZEB	Zero-Energy Building

CHAPTER 1

INTRODUCTION

"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

– Sherlock Holmes (Sir Arthur Conan Doyle)

Data-driven methods have become essential to solving many of the world's pressing problems in healthcare, climate change, military applications, transportation, and energy delivery. There is rising interest in the use of machine learning in such applications because machine learning makes it possible to infer models of systems from empirical evidence. Those models can be used to make predictions of future events and can characterize system behavior. Although the methods used to make those predictions and characterizations are often suitable for a wide range of applications, it helps to understand the nuances of each specific application so that the methods can be tuned to perform well for that application. This dissertation is concerned with the use of data-driven methods for applications related to the electric power grid.

The modern power grid, known as the *smart grid*, uses computer communication networks to improve efficiency by transporting control and monitoring messages between devices. Broadly speaking, those messages aid in monitoring the health of the grid and ensuring that electricity generation meets the constantly changing demand in a manner that minimizes costs to the grid stakeholders. In buildings, sometimes called *nanogrids*, communications between electrical appliances and centralized controls allow for more efficient electricity use. At the core of these technological advances are data, and it is vital to protect the integrity of those data because compromised data could undermine the efficiency benefits of modernization. Furthermore, such compromise could have both economic consequences, such as power theft, and safety-critical consequences, such as blackouts. The main contributions of the work presented in this dissertation are in the use of data to model generation and consumption, for addressing both the aforementioned concerns of grid efficiency and security.

Power grids are complex systems, and the U.S. power grid is considered to be the largest interconnected machine in the world [1]. Broadly speaking, all systems can be analyzed using two orthogonal modeling approaches: theoretical and empirical. Theoretical models may be either stochastic, wherein assumptions are made about probability distributions of the systems' state transitions, or deterministic, wherein well-defined rules govern the systems' operations. Those well-defined rules may be based on physics, and encoded in mathematical representations, such as differential equations.

Empirical models, unlike theoretical models, are based on data obtained from implementations of systems. By design, they are more realistic than theoretical models because they are based on real data, which serve as evidence of actual system behaviors. They do, however, need to be *retrained* for every system because a model constructed from the data of one system may not generalize well to other, similar systems. Furthermore, empirical models suffer from three main limitations. First, there needs to exist an implementation of a system in order to collect data from it for analysis. Therefore, comparisons of system designs are not possible unless the designs have been implemented. Second, instrumenting systems to collect data can be expensive. Often, that data is confidential and access to that data is restricted. Third, the computational requirement for processing that data can be large, and that increases the overall cost of using empirical models.

Sanders and others proposed the use of theoretical, stochastic models in the 1980s because such models do not suffer from the three aforementioned limitations of empirical models [2]. However, in the 30 years since those limitations were pointed out, a lot has changed, and empirical models have now become more popular than theoretical models. The recent increase in the use of empirical models has been driven largely by the reduced cost of both sensors (to perform measurements) and computational resources (to process the data). Furthermore, datasets that contain measurements of systems have become increasingly available in both academia and industry. Although empirical models require the system to be implemented, they are ideal for enabling better use of existing system capabilities to achieve the system operator's goals. For the analysis of complex systems, such as power grids, which have thousands of interacting components, empirical models are preferred because they scale better than stochastic, state-based models.

In this dissertation, we use empirical models to capture behavioral patterns in generation and consumption, which in turn aid in designing approaches to improve efficiency and security in smart grids. Before we can state the research objectives and contributions of this dissertation precisely, we need to provide a general characterization of different data-driven methods and explain how to apply them to smart grid data. We review those two topics in Sections 1.1 and 1.2, respectively. Then, we state the problems addressed in Section 1.3, the research objectives in addressing those problems in Section 1.4, and the limitations of the state of the art in addressing those objectives in Section 1.5. Our main contributions address those objectives, and we present them in Section 1.6, along with an explanation of how those contributions are organized into the dissertation chapters.

1.1 Characterization of Relevant Data-Driven Methods

We now provide an overview of data-driven methods, based on machine learning and statistics, that are relevant to this dissertation. Machine learning is one of the fastest growing areas of research at the time of this writing [3]. It uses empirical models, based on statistics and linear algebra, to provide insights about the underlying data. Its applications are broad-ranging and include computer vision, natural language processing, information retrieval, genomics, and analyses of various time-series data. This dissertation is concerned only with time-series data, and, in this section, we provide a general characterization of machine learning and statistical methods for analyzing such data.

As the name suggests, time-series data is collected from one or many sources and indexed by timestamps. The timestamps may either be confined to limited durations, such as hours in a day, or extend into the future indefinitely. A single time-series always has the same type of data; in the case of physical measurements, the data have the same physical units. The data may be textual (as in the case of computer system logs) or numeric (as in the case of power measurements). Ultimately, machine learning methods require that all non-numeric data be encoded in a numeric format before they can be processed.

Hundreds of machine learning methods have been developed by both academics and researchers in industry to analyze data, and many of those methods apply to time-series data.

One of the most significant unsolved challenges in machine learning is the identification of optimal methods to model a given dataset. For most applications, optimal methods for data analysis are not known; researchers evaluate multiple methods that they believe are appropriate for a given dataset and recommend the method that produces the best model, as evaluated on a specified performance metric. Each method is characterized by parameters and hyperparameters that can be tuned for improved performance, and that tuning is specific to a particular application. For example, in the application of machine learning to computer vision, the annual ImageNet Large Scale Visual Recognition Challenge defines problems that are tackled by research teams from around the world [4]. Although the problems in those challenges are well-formulated, researchers have been unable to identify optimal approaches for solving them. Instead, they continue to resort to tweaking the parameters and hyperparameters of existing machine learning methods to find an approach that surpasses all other approaches, as per the performance metrics defined in the challenges.

Although we are unable to identify optimal methods for modeling specific datasets, we can employ a well-reasoned approach for identifying methods that are suitable for those datasets. We propose a two-step approach. First, a preliminary analysis or observation of the data is performed through the use of visualization tools. Second, from that analysis, a hypothesis about the structure of the data is made. In the case of time-series data, the first level of characterizing a dataset is in recognizing whether or not the dataset has repeating patterns. Once that is determined, a deeper exploration of repeating or non-repeating patterns can be made to form a more detailed hypothesis about the dataset structure. Based on that detailed hypothesis, an appropriate statistical or machine learning approach can be adopted for analysis. The approach is illustrated in Fig. 1.1, wherein the arrows denote the progression of the decision-making process for identifying an appropriate set of methods for analysis.

As illustrated in Fig. 1.1, certain techniques are particularly suitable for datasets that have repeating patterns in them, while others are suitable for datasets that do not have repeating patterns in them. Going one step deeper in forming a hypothesis about the data, there are different ways by which repeating patterns can be characterized. The presence of a low-rank approximation implies that a trend repeats itself in a linearly dependent fashion. In such scenarios, dimensionality reduction techniques like principal component analysis

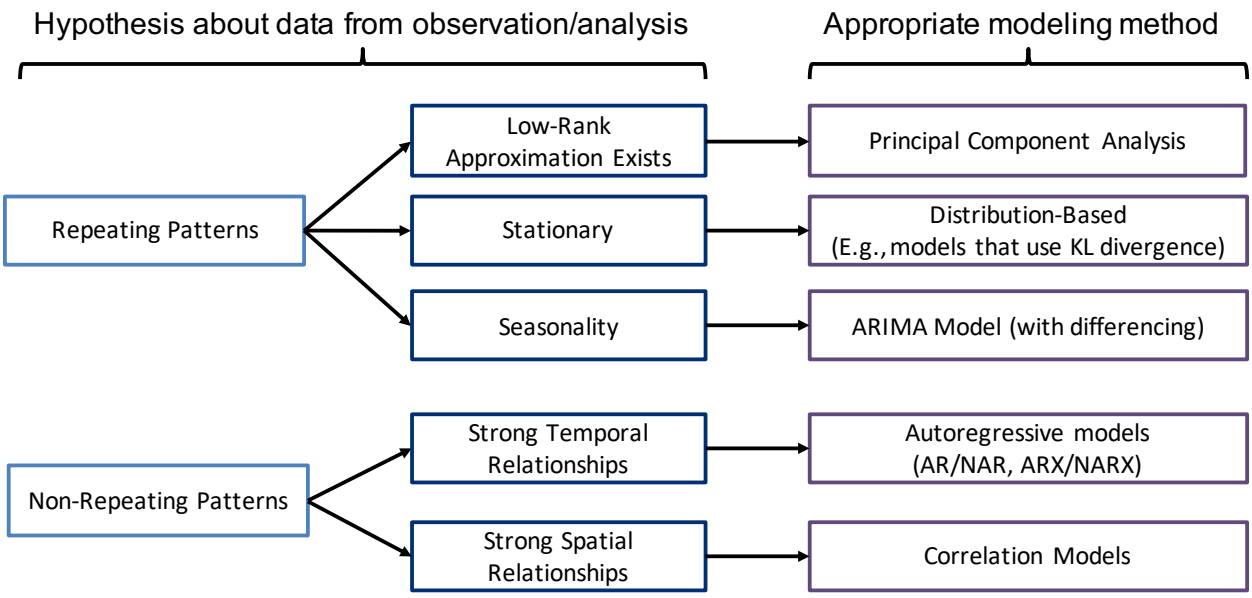


Figure 1.1: Identifying which modeling techniques are suitable for different types of data.

(PCA) work well. When there are cyclic patterns in the data, autoregressive integrated moving average (ARIMA) models are suitable for reducing model complexity by reducing the cyclic dependencies on data points from the distant past. When the data are stationary, the shape of their probability density function (PDF) repeats across different time windows and Kullback-Leiber (KL) divergence is a good measure to detect PDF changes for anomaly detection. The mathematical details of the aforementioned approaches will be discussed when they are applied, later in the dissertation.

Although the data may not have repeating patterns, there may be other characteristics that enable the creation of suitable models, such as spatial and temporal relationships between data points. Temporal relationships are useful in both prediction and anomaly detection applications. For prediction, data points from the past are used to predict data points in the future in autoregressive (AR) models. The AR models may be nonlinear (NAR) or they may take exogenous inputs from other data sources (in which case they are abbreviated ARX/NARX). In the absence of temporal relationships, spatial relationships could be exploited to cluster data points using distance metrics. For example, spatial density-based clustering approaches can be used for anomaly detection. Correlation models are useful when the relationship between two sources of data does not change over time.

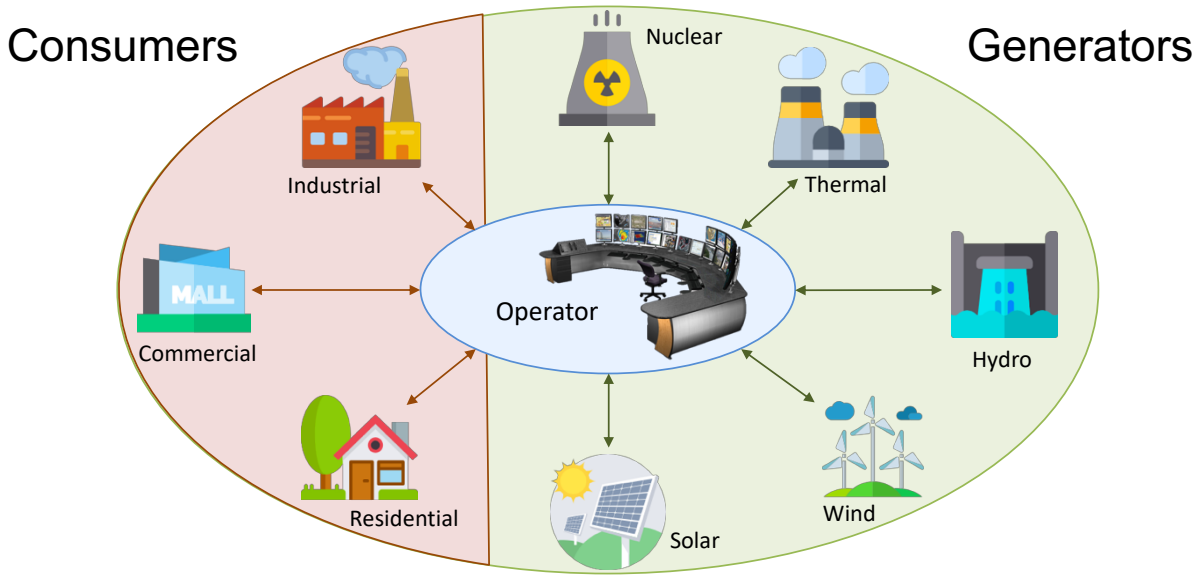


Figure 1.2: Consumers and generators in a smart grid communicate either directly or indirectly with operators.

1.2 Application of Data-Driven Methods to Smart Grid Data

Now that we have reviewed data-driven methods for time-series data in general, we will take a closer look at how those methods can be applied specifically to data collected from entities in smart grids. To understand that data, it is important to understand the entities themselves and how the data is collected from them. In this section, we describe the entities and characterize their data.

1.2.1 Overview of Smart Grid Entities

Broadly speaking, there are three entities in smart grids: generators, consumers, and system operators. As illustrated in Fig. 1.2, there are different types of generators and consumers, and they, either directly or indirectly, interact with a centralized system operator through computer communication networks. The role of the operator is to ensure that consumer demand is met by generator supply such that overall costs incurred are minimized.

Generators can be characterized based on whether or not they make use of turbines to convert mechanical energy to electrical energy (through electromagnetic induction). Almost all generators are driven by turbines, with solar power being a notable exception. The

turbines themselves are powered by resources, such as steam (produced by heat from burning coal or nuclear reactions), wind, or hydro. Renewable energy refers to energy that can be generated from resources that are not depleted after their use. Solar, wind, and hydro are the most common examples of renewable energy resources. The use of renewable energy resources is advantageous in that it does not lead to the depletion of natural resources, and it is often associated with lower carbon emissions in comparison to the emissions produced from the use of fossil fuels. The disadvantage in using renewable resources is that they are intermittent and their availability cannot be relied upon.

Electricity consumption happens at buildings, which can be classified into residential, commercial, and industrial. Electric vehicles can be charged at all those types of buildings. Commercial consumers are typically larger than residential consumers, and their electricity consumption patterns tend to be more regular. Industrial consumers are typically larger than commercial consumers and operate machinery that often require reactive power supply (for inductive loads) in addition to active power supply (for resistive loads). Typically, consumers receive electricity from generators through a series of conductors and voltage transformers. Some consumers produce electricity through the use of rooftop solar panels, and they are referred to as *prosumers*. Larger consumers may also have their own generation capabilities, which are typically used only when there is a need for backup power during an outage.

Most consumers consume electricity at low-voltage levels, and are connected to the low-voltage distribution grid. Large generation units also produce electricity at low-voltage levels, but that electricity is stepped-up to a higher voltage (to reduce transmission losses over long distances) and fed directly into a high-voltage transmission grid. The transmission grid feeds electricity into multiple distribution grids through transformers at substations that step-down the high-voltage electricity.

Consumption and generation levels are measured through supervisory control and data acquisition (SCADA) systems. Those systems are used for monitoring the health of the electric power grid and for ensuring that consumption and generation levels are properly tracked for energy market settlements. In most developed regions, consumers in the distribution grid are metered through an advanced metering infrastructure (AMI) operated by a local electric utility. Some meters, called *net meters*, can also be used to measure the generation from

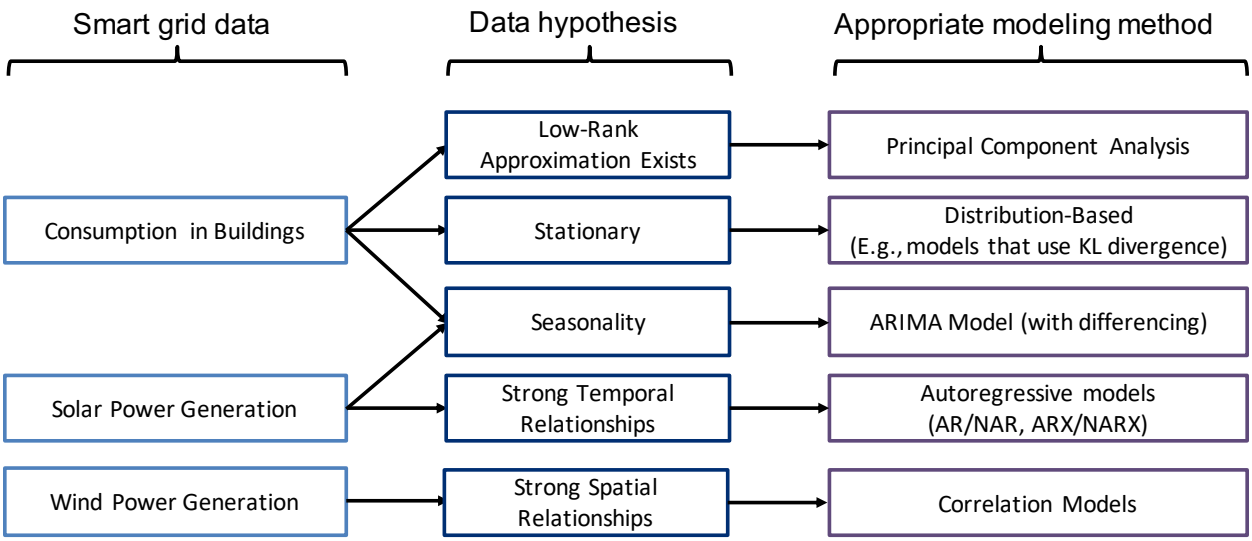


Figure 1.3: Assignment of data-driven methods for different smart grid data sources.

prosumers' rooftop solar panels. Large generators are metered through SCADA systems operated by the centralized system operator.

1.2.2 Characterization of Smart Grid Data

The general framework for choosing methods for modeling different types of data, which was illustrated in Fig. 1.1, is applied to smart grid data in Fig. 1.3. That data is collected from SCADA systems, and we consider data from consumers, wind power generators, and solar power generators. Hypotheses about that data can be made from preliminary analyses and data visualizations. We will provide examples of such analyses and visualizations later in the dissertation.

We now explain the characterization of the data for consumption in buildings, solar power, and wind power, as illustrated in Fig. 1.3. Electricity consumption data from buildings typically have repeating patterns, wherein consumption repeats on a weekly basis. All three types of repeating patterns illustrated in Fig. 1.1 may be observed in the data.

Similar to electricity consumption in buildings, solar generation data also exhibits cyclic patterns as a direct consequence of diurnal solar irradiance patterns. Apart from cyclic patterns, the other types of repeating patterns may not be as clear as they are with electricity

consumption. That is because solar generation varies significantly with time of the year (which affects the amount of daylight in a day) and weather conditions, such as cloud cover. Spatial relationships may best describe solar generation patterns. The reason is that nearby solar generators are more likely to be simultaneously subjected to the same variations in both the amount of daylight in a day and the cloud cover. Therefore, cross-correlation models are a good way to analyze the solar data. Those models do not apply well in the case of electricity consumption because consumption schedules can differ dramatically between neighboring buildings.

Wind power data are characterized by non-repeating patterns. Predictions made for future values of wind power are most influenced by current and recent-past values of wind power. Therefore, strong temporal relationships can be exploited to provide better prediction performance through the use of autoregressive (AR) models that may have exogenous inputs (in which case they are abbreviated as ARX). Wind power can also be modeled using wind speed data using a nonlinear relationship called the *power curve*. If the power curve model is unknown, it can be estimated using neural networks.

As argued at the start of the chapter, one limitation of data-driven approaches is that they cannot be used to model either systems that have not been instrumented or systems for which the data is not accessible. In such cases, simulations based on theoretical, physics-based models can be used. In the case of smart grids, for example, the EnergyPlus simulator can be used to model the air-conditioning systems of commercial buildings and generate synthetic measurements of those systems' electricity consumption. Those measurements can be generated for different temperature set-points and for different external weather conditions. As another example, PowerWorld or the Real-time Digital Simulator (RTDS) can be used to study synthetic data from synchronous generators and simulate the effects of cyber-attacks that can induce power outages through the manipulation of those generators. In that example, the attacks can be evaluated in a safe simulation environment without any hazardous real-world consequences.

1.3 Problem Statement: The Good, The Bad, and The Ugly

Now that we have provided background on data-driven methods appropriate for smart grid data, we proceed to define the problems that we will tackle using those methods in this dissertation. We address three problems that were brought about by the advent of communication technologies that enable smart grids; each of those problems has an associated cost of *several billions of dollars*.

1.3.1 The Good: Resource Utilization

As described in the start of the chapter, the communication capabilities that were introduced to make power grids *smart*, have enabled improved ease of management of grid assets. That in turn has enabled improved utilization of clean energy resources, better detection and isolation of non-malicious faults, and improved efficiency of monitoring the health of the grid.

Although smart grids enable better resource utilization, clean energy resources are being under-utilized because the availability of those resources is uncertain. For example, 21% of wind power is curtailed in China because wind power providers commit to providing less power to meet the consumption load than they actually produce; committing less power and providing more power leads to curtailments. The associated cost was over $1.2 billion from 2010 to 2016 [5]. In Germany alone, wind power curtailment costs were $0.5 billion in 2015 [6]. The good consequence of having smart grids is that increased connectivity and data sharing can enable better utilization of clean energy resources and reduce those curtailment costs.

1.3.2 The Bad: Fraud

Although assets such as smart devices and networking infrastructures enable the benefits of smart grids, the inadequate effort to secure those assets has made it possible for them to be compromised for adversarial gains. For example, data can be compromised by consumers and generators for fraudulent monetary gains. Fraud costs utilities billions of dollars every

year. In India alone, fraud due to electricity theft costs over $13 billion annually [7]. Smart meters are not prevalent in India, but they are being adopted in developed countries like Canada specifically with the intention to curb meter fraud [8]. However, smart meters are not a sufficient means for detecting such fraud; real incidents have been reported wherein smart meters have been compromised for fraudulent monetary gains [9]. The cost associated with meter fraud through the hacking of smart meters was estimated at $400 million annually in Puerto Rico in 2010 [10].

Although meter fraud costs utilities billions of dollars, the general public is not directly affected; the utility losses may be indirectly recovered through fees levied on consumers. Even so, the impact is purely financial.

1.3.3 The Ugly: Cyber-Outages

Certain vulnerabilities in smart grid systems can be exploited by cyber-adversaries to cause power outages, which can disrupt daily life and the economy. The increase in smart devices and communication-driven controls brought about by smart grids has increased the attack surface, and made it possible for adversaries to exploit more such vulnerabilities.

As part of the Aurora generator test in 2007, researchers at Idaho National Laboratories demonstrated that supervisory control and data-acquisition (SCADA) systems in generation controls can be compromised by a remote adversary [11]. In 2009, Stuxnet became the first malware to infect an industrial control system (in Iran) and cause it physical damage [12]. In 2016, Crash Override became the first malware to cause a power outage (in Ukraine) [13]. The ugly consequence of having smart grids with increased communication capabilities is that it has become easier for remote adversaries to cause power outages and damage to power system assets.

The power grid is essential to daily life, supporting other critical infrastructures like water, gas, transportation, and defense infrastructures. For that reason, and because large scale power outages can cost the U.S. economy hundreds of billions of dollars, protecting the power grid for cyber attacks that can cause power outages is a matter of U.S. national security [14].

1.4 Research Objectives

We address the problems relating to the good, the bad, and the ugly consequences of smart grids stated in the previous section through three research objectives. We state the objectives below before we describe each one of them in detail.

1. To improve the utilization of unpredictable clean energy resources and reduce curtailment costs.

2. To detect and mitigate meter fraud, which may be committed by both consumers and generators.

3. To understand, prevent, and mitigate attacks on generation controls that can cause outages by undermining the resilience of the grid to faults.

The first objective relates to improving *resource utilization*. Wind power and demand response are both sources of clean energy. While wind power is a form of clean energy generation, demand response through demand reduction is often thought of as the cleanest energy resource. The reason is that a reduction in load not only makes power available for other loads, but also decreases carbon emissions. Despite their benefits, both wind power and demand response are severely under-utilized because their availability is too uncertain to be relied upon by operators. For example, 21% of wind power was curtailed in China in 2017 because operators could not quantify that uncertainty [15]. Because of that, they consistently promised less wind power than they could generate. Our objective is to use data to quantify and reduce that uncertainty to facilitate better utilization by operators.

The second objective relates to improving *fraud detection*. The benefits of data-driven approaches to improve resource utilization can be undermined if the data have been compromised. In particular, fraudulent consumption and generation undermine the economic value of proper resource utilization. Such fraud can lead to losses of hundreds of millions of dollars for utilities in a single year [10]. Although the cases of fraud that we know of only deal with fraudulent consumption, it is conceivable that generators may as also resort to committing fraud to increase their income from generation. Our objective is to miti-

gate fraud by designing detectors that are based on data-driven models of generation and consumption.

The third, and final, objective relates to improving *cyber-resilience*. Unlike meter fraud, for which the consequence is purely monetary, attacks on resilience that seek to cause power outages can disrupt both the livelihoods of people and the economy. For example, a malware known as Crash Override created an outage in Ukraine in 2015 through crafting of malicious commands that could disconnect consumers from generators [13]. In this dissertation, we seek to understand similar attacks that aim at causing outages, by inhibiting the resilience of the grid to faults, through the compromise of generation controls. Our objective is to propose methods to prevent, detect, and respond to such attacks.

Apart from the three aforementioned technical objectives, the dissertation also aims at presenting applied research that can be transitioned to industry.

1.5 Limitations of the State of the Art

In this section, we describe the limitations of the state of the art in addressing the three research objectives stated in Section 1.4; in subsequent chapters, we will provide a more comprehensive and detailed description of the literature with citations to related work.

In the context of wind power utilization, the literature falls short in quantifying and minimizing the uncertainty associated with wind power prediction. Instead, it focuses entirely on methods that reduce the mean of the prediction error. There is a need to reduce not only the mean of the prediction errors, but also the standard deviation. Doing so increases the operator's confidence in wind power prediction so that it can be better utilized. Similar to wind power utilization, in the context of demand response utilization, the literature has made no effort to quantify uncertainty of demand response capacity.

The literature on meter fraud detection does not explore the space of attack vectors that can be used to accomplish such fraud. It is limited in that only certain types of fraud are addressed. For example, different attacks apply under different electricity pricing schemes, such as flat pricing or time-of-use pricing. A formal approach to identifying attacks is missing, and therefore several attacks were not identified in the literature. The identification

of attacks is important so that appropriate detection methods can be designed for those attacks. Although detectors are proposed in the literature, an analysis of worst-scenarios for those detectors was not presented. The possibility of fraud committed by generators was also not addressed, though generators have a real incentive to commit fraud. A data-driven analysis of detectors for both consumption and generation fraud is missing in the literature.

In the context of cyber-resilience, the literature addresses attacks on secondary frequency response (SFR) mechanisms in power grids, which provide resilience when there is a sudden loss of generation due to a fault. However, attacks on primary frequency response (PFR) mechanisms, which take effect before SFR mechanisms, are not studied. The study of possible attacks on PFR is important because those attacks can lead to outages before SFR takes effect. With the increase in wind power penetration, system operators in many countries have begun to use wind power for PFR. We show that although wind turbines are susceptible to the same attacks on PFR, the attacks can be detected by using empirical models that map wind speed measurements to expected wind power generation.

1.6 Summary of Main Contributions

To achieve the research goals outlined in Section 1.4, we use machine learning and statistical methods to construct empirical models of electric power generation and consumption using real data obtained from various subsystems in the grid. Through the work presented in the dissertation, we show that such a data-driven approach outperforms heuristic approaches. That forms the basis of our thesis statement, which follows.

Thesis Statement: Empirical models of generation and consumption, constructed using machine learning and statistical methods, improve resource utilization, fraud detection, and cyber-resilience in smart grids.

Resource utilization, fraud detection, and cyber-resilience are the primary themes of this dissertation, and our contributions are organized into chapters according to those themes. The organization of chapters is summarized in Table 1.1. As shown in the table, we address resource utilization and fraud detection for generation and consumption in separate chapters. Although we study cyber-resilience in the context of attacks on generation controls, those attacks also have an impact on consumers because they can cause power outages. Therefore, our work on all three primary themes addresses issues related to both generation and consumption. Next, we provide a brief summary of our contributions for each of those three themes.

Table 1.1: Organization of topics covered

Ch.	**Primary Theme**	**Smart Grid Topic**
2	Resource Utilization	Generation: Wind Power
3	Resource Utilization	Consumption: Commercial Buildings
4–6	Fraud Detection	Consumption: Residential and Commercial Buildings
6	Fraud Detection	Generation: Wind & Solar Power
7	Cyber-Resilience	Generation: Synchronous & Wind Power

1.6.1 Resource Utilization: Chapters 2 & 3

In Chapter 2, we use wind speed and wind power data obtained from sensors on wind turbines to quantify uncertainty and improve utilization of wind power. We improved the wind power prediction accuracy for up to a 6-hour look-ahead by using auto-regressive models with exogenous inputs from hyper-local weather forecasting data. In doing so, we used the framework presented in Fig. 1.3; we exploited the strong temporal relationships in wind power data and used neural networks to estimate the relationship between wind speed and wind power. We also augmented the improvement in prediction accuracy by decreasing the uncertainty associated with the prediction. Together, those improvements allow utilities

to make better use of wind power generation by allowing them to better predict how much generation can be expected in the near-term.

In Chapter 3, we use data obtained from sensors that we deployed in commercial building spaces to quantify uncertainty and improve the utilization of demand reduction as an energy resource. Using estimates of occupancy, we quantify the utilization and wastage of electricity at different times and for different loads. Furthermore, the ability of buildings to provide demand reduction as a resource is constrained by occupant comfort requirements. Finally, we use a probabilistic approach to quantify uncertainty associated with demand reduction at different times of the day based on the aforementioned occupancy and comfort metrics. That quantification of uncertainty informs operators, so that they can call on select buildings to provide the required demand reduction with minimum uncertainty.

1.6.2 Fraud Detection: Chapters 4–6

Chapters 4–6 deal with the economic consequences of attacks on the integrity of smart grid data. In particular, they are concerned with modifications to consumption and generation readings that lead to fraudulent monetary gains.

Chapter 4 provides a framework for identifying different classes of attacks that can cause fraudulent monetary gains. It considers the environmental factors in which the attacker is operating, such as pricing systems, the existence of state-of-practice detection mechanisms, and the availability of automated demand response. It obtains the necessary conditions required for an attack to be successful in each of those environments. The framework is used in Chapters 5 and 6 for the evaluation of specific attack detection algorithms.

Chapter 5 focuses on the detection of electricity theft accomplished through the compromise of meter data. It explores the use of data-driven detection methods suitable for all three types of repeating data described in Fig. 1.3. The detection methods are compared based on how much money can be gained by an attacker by circumventing attacks proposed in related work. The methods are also compared in terms of their receiver operating characteristics.

For each detection mechanism proposed in Chapter 5, an optimal attack that circumvents that mechanism is presented in Chapter 6. The optimal attack maximizes the fraudulent

monetary gain while circumventing detection. Data-driven methods from Chapter 5 are augmented with new methods that mitigate fraud for solar and wind generators in Chapter 6. Appropriate modeling techniques for wind and solar data are identified using the framework illustrated in Fig. 1.3. Weather data are used for mitigating wind generation fraud, which is significantly harder to detect than solar generation fraud because wind is more erratic. For each new detection mechanism, an optimal attack that circumvents that mechanism is presented. The detection methods are also compared in terms of their receiver operating characteristics. In addition, this chapter presents a study of how quickly generators can recover the capital costs of their generation facility by committing fraud.

1.6.3 Cyber-Resilience: Chapter 7

We use the PowerWorld simulator to demonstrate how attacks that change turbine governor control settings can cause loss of resilience. Those settings determine how much a generator would increase its generation to compensate for loss of power due to faults elsewhere in the grid. When there is insufficient compensation for lost power, the system frequency drops, and that drop triggers protective mechanisms that perform under-frequency load-shedding. That load-shedding can cause service disruption, and can be catastrophic. We evaluate the impact of various attack parameters to determine when the system is at risk of outages. The evaluation is performed for synchronous generators and wind turbine generators. Data-driven detection strategies are proposed.

We conclude the dissertation and describe its impact in Chapter 8. We include some additional work on the dissertation topics that are not directly relevant to the research objectives in the Appendix. The contributions have been included in peer-reviewed publications that are listed in Appendix A. Having summarized our main contributions and laid out the organization of the dissertation, we now proceed to expand on each contribution in the chapters that follow.

CHAPTER 2

DATA-DRIVEN WIND POWER INTEGRATION

"Errors using inadequate data are much less than those using no data at all."

– Charles Babbage

Globally, wind capacity increased by 17% from 2014 to 2015 [16]. In the U.S., 41% of capacity additions in 2015 came from wind power [17], which was more than that from any other energy source. Wind power is playing a major role in meeting electricity demand in an increasing number of countries, including Denmark (42% of demand in 2015), Germany (more than 60% in four states), and Uruguay (15.5%) [16]. However, the limited predictability of this weather-dependent energy source has become a growing concern in power grids, and sometimes forces utilities to curtail wind generation. In China, for example, 21% of wind power was curtailed on average in 2017 [15]. Our contributions in this chapter aim at providing utilities with better wind prediction capabilities, allowing them to decrease such curtailments and improve the utilization of wind power generation potential. The solution was operationalized for a utility in Vermont, but the techniques are applicable worldwide.

Wind power scheduling is typically performed through a bidding process wherein the supplier commits to providing a certain amount of power at specified times of the upcoming day. In day-ahead scheduling, that commitment is made for the next day, while in spot (real-time) markets, the day-ahead agreement is adjusted for the immediate future (on the order of 10 minutes to 6 hours ahead). Accurate and precise predictions of how much power can be generated at a certain time are essential if a supplier is to place a reliable bid. More generally, accurate and precise forecasts are in the economic interest of wind farm owners, independent power producers, utilities, transmission system operators, and their consumers, and provide environmental benefits. To aid in prediction, many related efforts have addressed wind power forecasting for the day ahead [18,19]. Optimal bidding strategies for wind power in the face of uncertainty were presented in [20].

Improvements in wind forecasting not only assist in electricity market bidding processes,

but also make it easier and more economical to schedule generation in microgrids [21, 22]. They also facilitate dynamic load scheduling [23, 24].

2.1 Summary of Contributions

In this chapter, we predict wind power with AutoRegressive eXogenous (ARX) and nonlinear ARX (NARX) models that use hyperlocal wind forecasts (with a spatial granularity of less than 2 miles) as exogenous inputs from a customized Weather Research and Forecasting (WRF) model similar to that in [25]. In doing so, we make three main contributions. First, by comparing results for different model inputs and outputs, neural network parameters, power curve fits, and training data sizes, we show how to fine-tune the model architecture such that accuracy is maximized. Second, we present an extensive evaluation of the models, and show that our approach reduces the prediction error to 2.11% when looking 10 minutes ahead, and to 14.25% when looking 6 hours ahead, for a wind farm of 21 turbines. Third, we make comparisons with approaches in related work to show that our approach improves both prediction accuracy and precision; it reduces the uncertainty (confidence interval width) associated with that prediction by over 15%. In summary, utilities can use our approach to make improved short-term predictions and mitigate the cost associated with dispatching too much (or the risk associated with dispatching too little) wind power.

In addition to the aforementioned contributions, we provide useful insights obtained from real data. When the inputs and outputs of the prediction model are both in the same units (wind speed or wind power), we show that nonlinear models do not fit our data any better than linear models would, suggesting that linear models are sufficiently accurate despite their simplicity. When converting from wind speed inputs to power outputs, however, we show that nonlinear models have a slight advantage over linear models. Those insights challenge the current hype around the use of deep learning for prediction.

This chapter is organized as follows. We describe the dataset we used in the study in Section 2.2. We formulate the wind power prediction model and describe specific variations of that model in Section 2.3. Two of those variations require the mapping of wind speed to wind power using a power curve, and we present a power curve estimation study in

Section 2.4. We evaluate the wind power prediction models in Section 2.5 and verify the results against theoretical expectations in Section 2.6. We discuss how our approach reduces the uncertainty associated with the predictions in Section 2.7, and how that helps improve resource utilization in Section 2.8. We present related work in Section 2.9 and conclude in Section 2.10.

2.2 Description of the Dataset

In this chapter we use real data that were obtained from 21 wind turbines in the state of Vermont. Wind speed measurements were taken every 10 minutes from anemometers installed on all the turbine units over a period of 403 days from May 7, 2015, to June 13, 2016. For each wind speed measurement, the data contained the power that was possible to generate from that wind speed.

Hyperlocal (within a 0.2–1.2 mile radius) wind velocity forecasts for each wind turbine were computed by a customized Weather Research and Forecasting model (WRF) similar to that in [25]. The forecasts were made for the day ahead with a time step size of 10 minutes. The WRF calculated the three velocity components at the turbine altitude of 84 meters, and we derived the wind speed by taking the L^2 norm of those components. The altitude is important to note because wind turbines are influenced by wind at the altitude of the turbine, which may differ substantially from the wind velocity at the ground level.

Assuming continuity between the data points, we filled gaps in the data set by linear interpolation before performing our analysis on it. Since only 2.6% of the data was missing, we did not investigate more sophisticated forms of interpolation. The inertia of meteorological systems justifies the aforementioned assumption of continuity between data points.

2.3 Prediction Models

We consider three prediction models that differ in their inputs. The output of all models is the predicted wind power. As part of the prediction procedure, we continuously collect measurements at a time resolution of $\Delta t = 10$ minutes. In our set-up, the hyperlocal

forecasting engine runs once every 24 hours, providing us with numerical weather forecasts at a time granularity of 10 minutes. Since we wish to make predictions 6 hours into the future, we require that the engine provide us with at least 30 hours of forecasts, so that a 6-hour-ahead forecast can be made at the end of each run. Each run of our WRF engine provides 72 hours of forecast data, and we overwrite older forecasts with the most recent forecasts because they are usually more accurate.

2.3.1 General Prediction Model

All three of the models that we present estimate a wind power prediction function f_P^*. At a given time t, f_P^* takes as input n_A past measurements (of wind speed or wind power, depending on the model, as discussed later in Section 2.3.2), and n_D hyperlocal (WRF) wind speed forecasts. f_P^* maps those inputs to h predictions of wind power for the next $h\Delta t$ min after time t. The n_A measurements were taken between times $t - n_A\Delta t$ and t. The n_D forecasts are taken between times $t - n_{DP}\Delta t$ and $t + n_{DF}\Delta t$, where $n_{DP} + n_{DF} = n_D$. In taking n_{DP} forecasts made for times before the current time t, we account for inaccuracies common to hyperlocal forecasting engines wherein the time for which a forecast was made is off by a fixed lag. For example, the forecasting engine may predict at the start of the day (12 A.M.) that the wind speeds from 10 A.M. to 4 P.M. will take on a certain set of values. However, the forecast may have been off by 30 minutes, so that set of values would actually be observed from 9:30 A.M. to 3:30 P.M. Since we seek to improve the h prediction values by using exogenous inputs from WRF, we set $n_{DF} = h$. We performed a sensitivity analysis on n_{DP} and found that increasing n_{DP} beyond n_{DF} provides the model with more data, but the accuracy did not improve, so we set $n_{DP} = n_{DF}$.

We estimate the function f_P^* by using a training set of actual measurement data and WRF forecasts. There are m samples in the training set, and each sample has $n = n_A + n_D$ features representing actual measurements and forecasts. Each sample is associated with a current time t and a known set (ground truth) of h future wind power values after t. As t slides through the dataset, a new sample of n features is obtained. In each new sample, the oldest value is removed, and the latest value is added to the n_A measurements and n_D forecasts in

the previous sample. Therefore, in any two consecutive samples, exactly $(n_A - 1) + (n_D - 1)$ values will coincide. The $m \times n$ training matrix is denoted by X, and each sample is denoted by X_i. The corresponding ground truth of predictions is an $m \times h$ matrix denoted by y, and each row is denoted by y_i.

We set up an optimization problem to estimate f_P^* from f_P candidates in a way that minimizes the error between the predictions $f_P(X_i)$ and the ground truth data y_i. There are multiple ways to define that error, so one could consider various objective functions, such as mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE). In this chapter, we chose MAE because it is commonly used in the industry and in the literature. The MAE averages the L^1-norm of the difference between the predicted values $f_P(X_i)$ and the ground truth values y_i in the training set. f_P^* is the optimal f_P that minimizes the MAE.

$$f_P^* = \arg\min_{f_P} \frac{1}{m} \sum_{i=1}^{m} ||f_P(X_i) - y_i||_1. \tag{2.1}$$

In our experiments, we chose $n_A = 6$ by trial and error. That means we "look back" 6×10 minutes $= 1$ hour to predict the output looking up to 6 hours ahead. Kusiak et al. [26] also looked back 6 time periods, and Blonbou [27] looked back 5 time periods.

L^1 Minimization Approach

In the case of an AR or ARX model, f_P becomes linear:

$$f_P^* = \arg\min_{f_P} \frac{1}{m} \sum_{i=1}^{m} ||X_i U - y_i||_1, \tag{2.2}$$

where U is an $n \times h$ matrix. The solution to the above minimization problem using linear programming is well-known (see [28]), and there is no need to use neural networks to solve it.

Feedforward Neural Network Approach

The L^1 minimization approach assumes that f_P is a linear function. It is possible to approximate f_P^* using neural networks, without placing any such linearity constraints on f_P in Eq. (2.1). The simplest neural network is a feedforward neural network with zero hidden layers and a linear activation function on the output. That type of network is essentially a linear regression model, and can be used to fit AR and ARX models. Nonlinear models such as NAR and NARX can be estimated with feedforward neural networks that contain one or more hidden layers. Each hidden layer contains a variable number of hidden neurons, which are intermediate states that represent combinations of their inputs. Those combinations are determined by the choice of *activation functions* on each layer, such as a linear, rectified linear unit (ReLU), sigmoid, or hyperbolic tangent function.

Feedforward neural networks map the inputs of the model to the outputs. Each input neuron of the network corresponds to a feature of the input, and each output neuron corresponds to a prediction at a specific horizon. If there are n input features, the neural network has n input neurons. There are h output neurons for the h horizons that we seek to predict. Between the input and output layers there exist $l \geq 0$ hidden layers. The width of the network, q, is the number of nodes in each hidden layer and is usually kept constant across layers for convenience. As a rule of thumb, $q \geq n$. The configuration is illustrated for $n = 9, h = 10, l = 1$, and $q \geq 10$ in Fig. 2.1. We used the Keras Python library [29] with a sequential architecture to implement feedforward neural networks. We choose the MAE as the objective function to minimize in the Keras model settings. In accordance with the state of practice, we configured each hidden layer with a ReLU activation function and the output layer with a linear activation function.

2.3.2 Specific Prediction Models

All three models are uniquely identified by their inputs. For notational convenience, let A and D denote *actual measurements* and *WRF exogenous forecasts*, respectively. Let the subscripts S and P refer to *wind speed* and *wind power*, respectively. For example, Model $A_P D_S$ refers to the model with *actual wind power measurements* and *WRF wind speed*

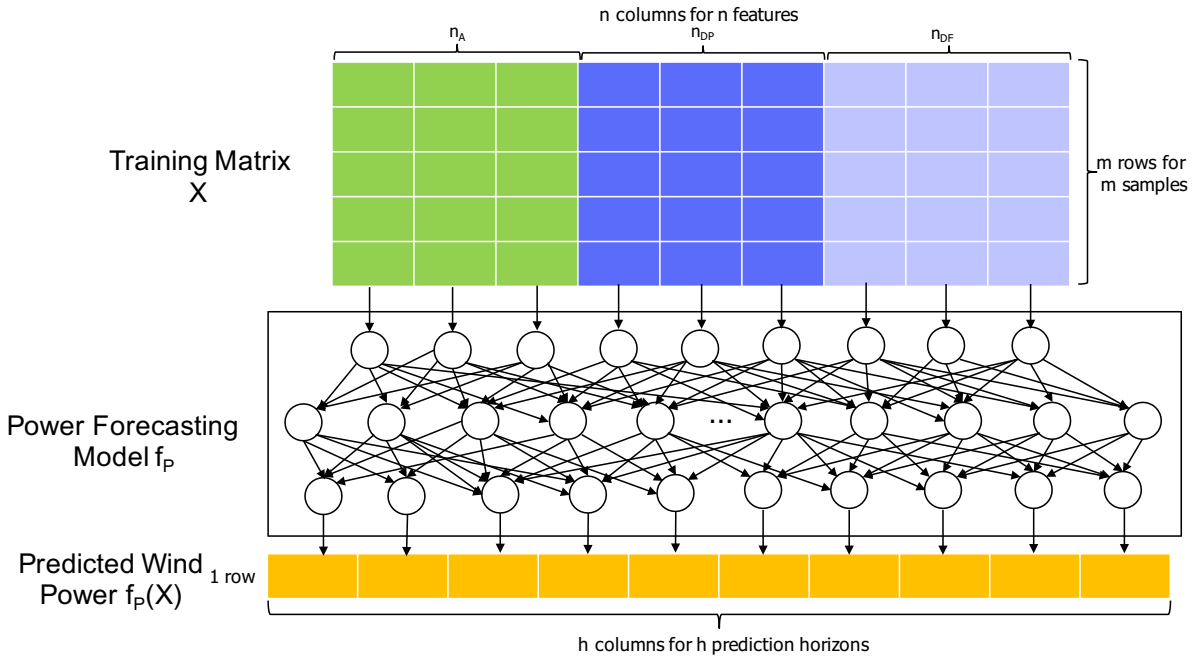


Figure 2.1: An example feedforward neural network with 1 hidden layer ($l = 1$).

forecasts as inputs.

In models $A_P D_S$ and $A_P D_P$, we assume that the electric utility does not have anemometer (wind speed) measurements available, but does have wind power measurements available (A_P). This assumption is validated from our own interactions with various electric utilities and knowledge of their sensing system capabilities. Some utilities have anemometer data, while others do not, and we seek to provide solutions for both situations.

Model $A_P D_P$

This model first maps the wind speed forecasts (D_S) to implied wind power forecasts (D_P) by using the *power curve* mapping function. The result of that mapping (D_P) is then combined with the actual power measurements from the turbine (A_P), so that the prediction function estimation occurs in the power domain. The corresponding training matrix is illustrated in Fig. 2.2. In this configuration, we find that a linear prediction model (described in Section 2.3.1) achieves the same accuracy as nonlinear models estimated using neural networks.

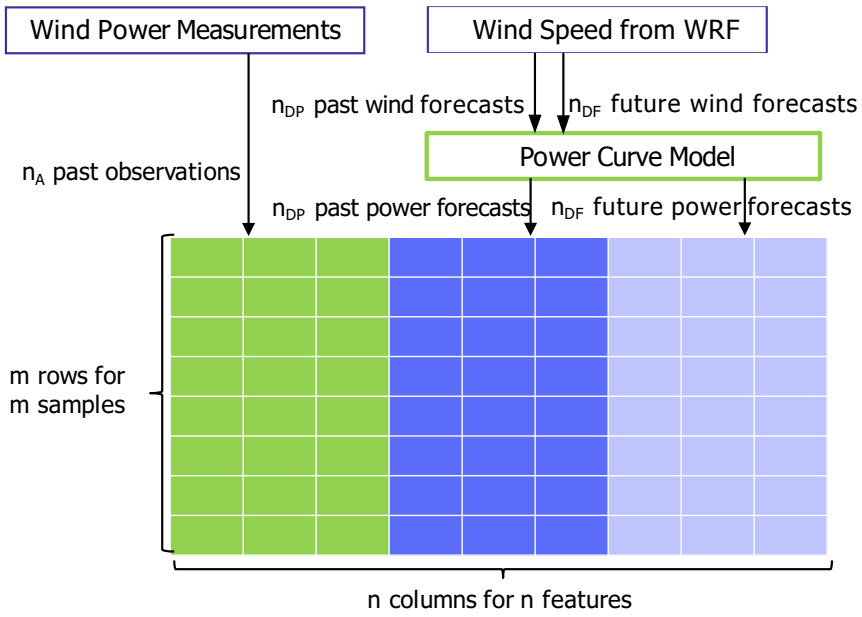


Figure 2.2: Illustration of training matrix X with m samples and n features for the $A_P D_P$ model. The n features contain n_A actual measurements and $n_{DP} + n_{DF}$ forecasts from WRF. The forecasts were mapped using a power curve model from wind speed to wind power.

Model $A_P D_S$

This model combines the wind speed forecasts (D_S) directly with the actual power measurements (A_P) to produce a mixed-domain input for the prediction function estimation. Here, the actual measurements are in the power domain, while the forecasts are in the wind speed domain. A power curve is not used in this configuration, and we find that nonlinear prediction models estimated by neural networks produce more accurate predictions than linear models do.

Model $A_S D_S$

This model combines the wind speed forecasts (D_S) directly with anemometer measurements from the turbine (A_S). In this approach, the optimization first outputs wind speed, so that the prediction function has inputs and outputs in the wind speed domain. Then, wind power is calculated using the power curve. In this configuration, we find that a linear prediction

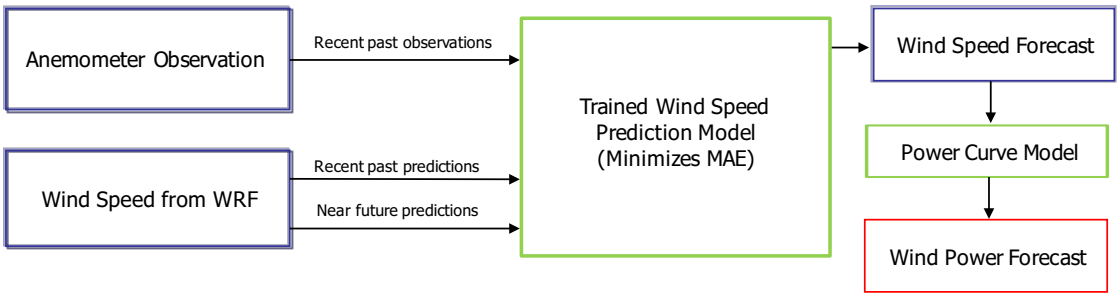


Figure 2.3: Indirect approach of $A_S D_S$: Predicting wind speed *before* converting to wind power.

model (described in Section 2.3.1) achieves the same accuracy as nonlinear models estimated by neural networks. Once the wind speed is predicted, wind power is obtained by using a trained power curve model that is nonlinear. The approach is illustrated in Fig. 2.3.

The authors of [26] also use an approach wherein they first predict wind speed and then convert to wind power. However, their model does not include exogenous inputs, and is therefore not accurate for prediction horizons beyond 10 minutes ahead.

2.4 Power Curve Estimation

The forecasted power for a particular turbine can in principle be obtained from the forecasted wind speed through use of the wind turbine manufacturer's power curve, which maps wind speed to possible power. That mapping should be performed during the construction of the training matrix, as illustrated in Fig. 2.2. Note that this mapping might be inaccurate in practice if the wind turbine operator poorly controls the angle of the turbine blades such that the possible power for that wind speed is not realized. For that reason, and because we do not have the manufacturer's power curve, we estimate the power curve from the data.

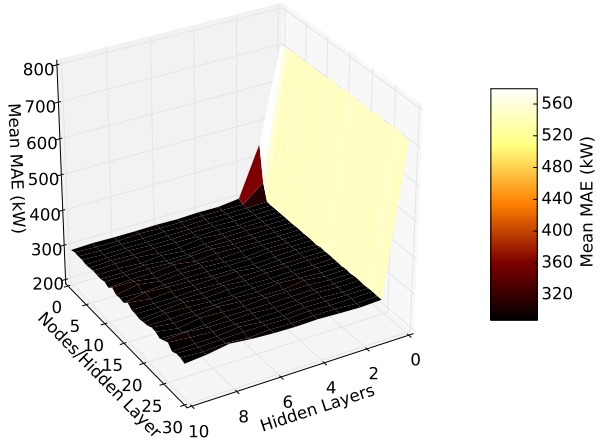
The mapping function from wind speed to wind power is approximated by a cubic function [30], but that function does not apply to wind speed values below the cut-in speed (below which the output is zero) and above the cut-out speed (above which the output saturates) of the turbine. Therefore, we propose an alternative approximation that uses a function estimated from a feedforward neural network, and we call that the *power curve*

neural network (*PCNN*). The PCNN is not to be confused with the neural network used to estimate the ARX/NARX prediction model illustrated in Fig. 2.1. It plays the role of the boxes corresponding to the power curve model in Fig. 2.2 and Fig. 2.3.

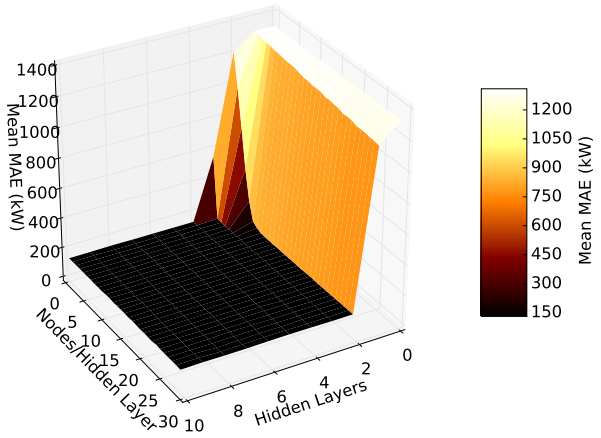
The PCNN has a single input neuron, and a single output neuron since the function maps scalar values. First, we need to choose the appropriate activation function on the input layer. A typical choice is the linear activation function, but we show from data-driven evaluations that linear activation functions perform worse than sigmoid and hyperbolic tangent activation functions. Second, choosing the number of hidden layers l and the number of nodes per hidden layer q can heavily influence the MAE and thus the prediction quality.

We trained the PCNN for various choices for the number of hidden layers l and nodes per hidden layer q, and illustrate their results in Fig. 2.4. In the three surface plots we applied linear, sigmoid, and tanh activation functions, respectively, on the first hidden layer and linear activation functions on all other layers. The sigmoid activation was also used in [27], and we found that both the sigmoid and tanh functions achieve the same level of accuracy in estimating the power curve. Figure 2.5 illustrates two power curves that use sigmoid activations with different depth (l) and width (q) parameters. When we used the linear activation function, the resulting minimum MAE was more than twice as large as the minimum MAE yielded when we used the sigmoid or tanh activation. The black regions in Fig. 2.4 are regions of MAE that are relatively stable with respect to the PCNN configurations (depth/width). For the sigmoid and tanh activation functions, this region is given by $l > 2$ and $q > 3$. Any choice of PCNN configuration in that region will produce an average MAE across all turbines of at most 138.11 kW and at least 129.36 kW. In other words, increasing the PCNN depth and width does not improve the prediction accuracy significantly as long as they are within that region of stability. The region is not convex, and there is no trend indicating that having more or fewer layers always achieves an advantage. We chose to work with $l = 6$ and $q = 22$ because that PCNN configuration produced the lowest MAE for both the sigmoid and tanh activations, when averaged across all 21 turbines.

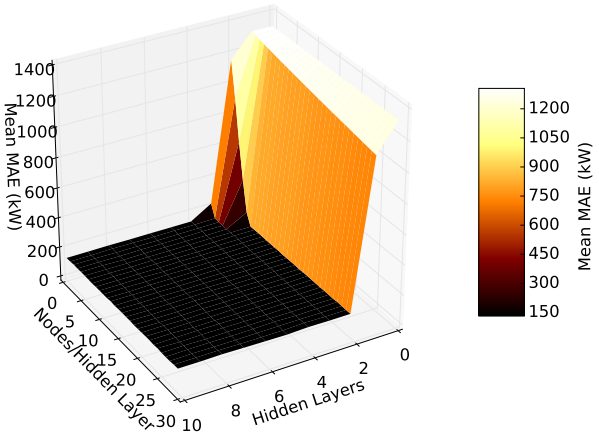
The key takeaway for practitioners is that there is no benefit to having a deeper/wider network once the configuration is operating in the region of stability, but there is significant value in finding the right activation function. The authors of [31] discuss various methods for



(a) Linear Activation
(minimum mean MAE is 284.82 kW)



(b) Sigmoid Activation
(minimum mean MAE is 129.36 kW)



(c) Tanh Activation
(minimum mean MAE is 129.92 kW)

Figure 2.4: MAEs of power curve neural networks (PCNNs), averaged across all turbines. Sigmoid and tanh activations achieve comparable accuracies, which are significantly greater than the linear activation accuracy.

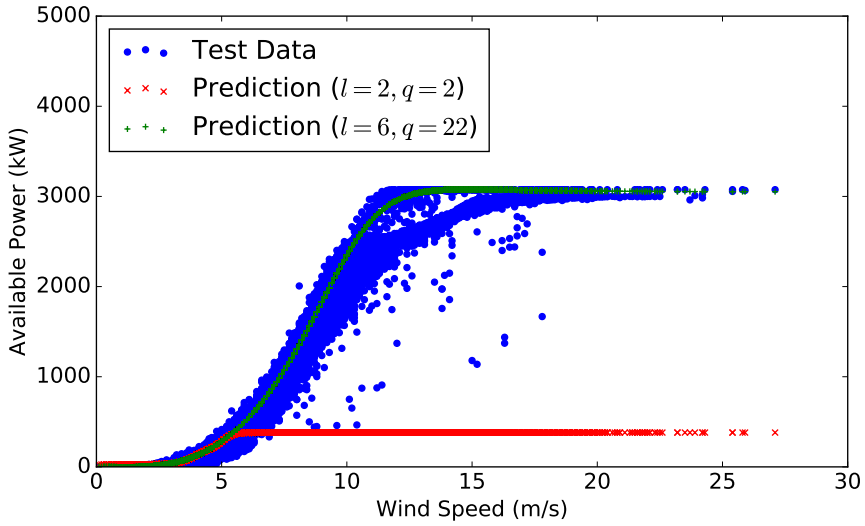


Figure 2.5: Two power curve fits, using two neural networks (sigmoid activation).

power curve estimation. An alternative approach to estimating wind power using k-nearest neighbors is proposed in [26]. Polynomial regression techniques are explored in [32]. Power curves were shown to be useful for generator fraud detection in [33].

2.5 Evaluation

In this section, we present a comprehensive evaluation of our approaches on the wind farm data.

2.5.1 Evaluation Metric

As explained in Section 2.3.1, we used the industry standard mean absolute error (MAE) to measure the prediction accuracy of our approaches. As that metric is in kW, as shown in Fig. 2.4, it is harder to interpret than a percentage would be. Therefore, we use another industry standard approach to normalize the MAE with the capacity of the turbine to express it as a percentage. The normalized MAE (NMAE) is defined as follows.

$$\text{NMAE} = \frac{MAE}{r}, \tag{2.3}$$

where r is the wind turbine rating if the MAE is computed for each wind turbine, or the aggregate rating of the wind farm if the MAE is computed for the wind farm as a whole. In our homogeneous data set of 21 wind turbines, all turbines had a rating of 3.07 MW, giving the wind farm a capacity of 64.57 MW.

Note that NMAE is different from MAPE in that MAPE normalizes each prediction error with the actual value, while NMAE normalizes it with the rating (or maximum possible value). MAPE is an unpopular metric for prediction because it produces abnormally large values (or divide-by-zero errors) when the actual value is near zero. We express NMAE as a percentage by multiplying it by 100%.

2.5.2 Cross-validation

We performed tenfold cross-validation to ensure that our results were not biased because of specific choices of training and test sets. For training, we used 20 days of data (2880 data points), of which 2 days (10%) were used for validation. We tested the model on 1 day of data (144 data points). The ten cross-validation groupings were such that each group contained 21 consecutive days of data for training and testing. Also, the groups were evenly spread throughout the 403-day period of the dataset, so the bias caused by the time of the year is accounted for.

We use η to denote the set of cross-validation groups and τ to denote the set of turbines. The test set is indexed by the times for which the predictions were made, and that set of times is denoted by ρ. Each sample in the test set is denoted by $X_{i,j,k}$, where $i \in \eta, j \in \tau, k \in \rho$, and the corresponding ground truth value is $y_{i,j,k}$. Our NMAE results are averaged over the ten cross-validation groups as follows.

$$\text{Average NMAE} = \frac{1}{|\eta|} \sum_{i \in \eta} \frac{1}{|\tau|} \sum_{j \in \tau} \frac{1}{|\rho|} \sum_{k \in \rho} \frac{|f_p^*(X_{i,j,k}) - y_{i,j,k}|}{r}, \tag{2.4}$$

where $|.|$ denotes the cardinality of the sets. In our experiments, $|\eta| = 10$, $|\tau| = 21$, and $|\rho| = 144$. In performing three stages of averaging, we were able to summarize the large number of prediction results ($10 \times 21 \times 144 = 30,240$ per horizon to be exact) that we

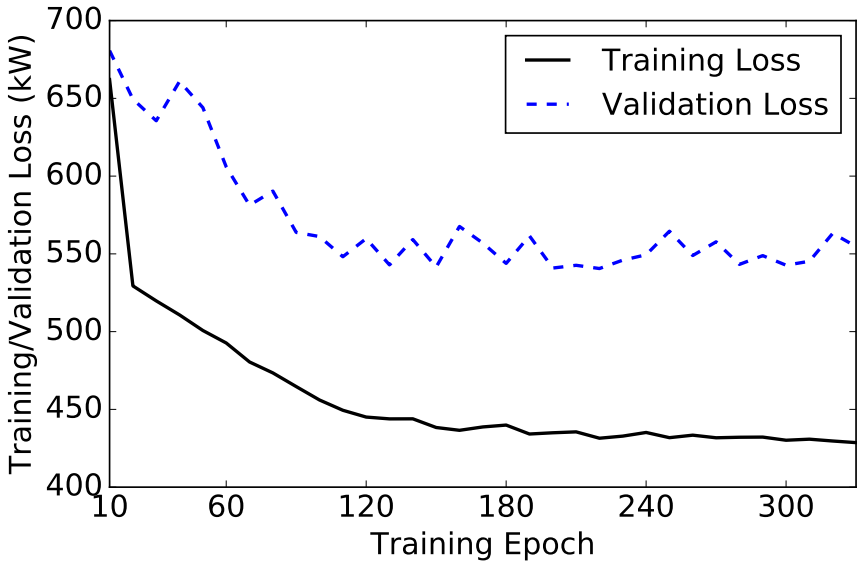


Figure 2.6: Training and validation loss (MAE).

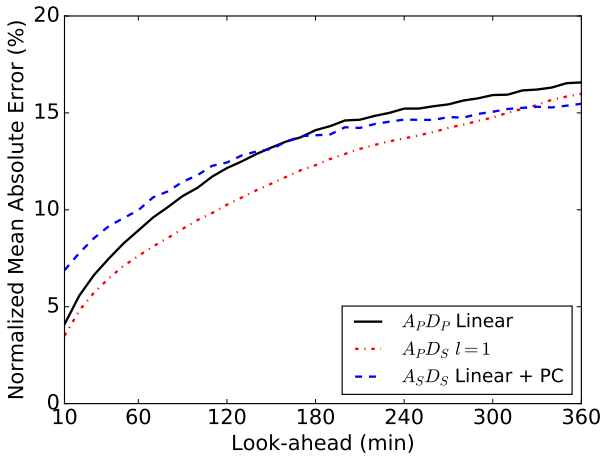
obtained for each of the approaches that were evaluated. That averaging, however, resulted in loss of information regarding the spread of the NMAE values in the three different groups. In Section 2.5.6, we discuss the spread of the NMAE across the sets ρ and η. In Section 2.5.7, we discuss the spread of the NMAE across the group of turbines τ.

2.5.3 Training Approach

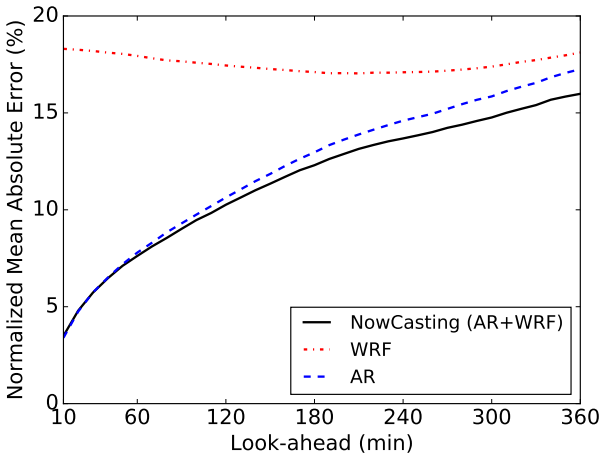
We trained our neural networks using the backpropagation algorithm [34] to minimize the loss function (MAE). We ensured that the training automatically stopped when the validation error ceased to decrease, or started to increase after 10 epochs. That is a recommended approach to preventing overfitting of the training set; an example is illustrated in Fig. 2.6. As a result of that approach, different neural network configurations ran for different numbers of epochs.

2.5.4 Results on Model Accuracy

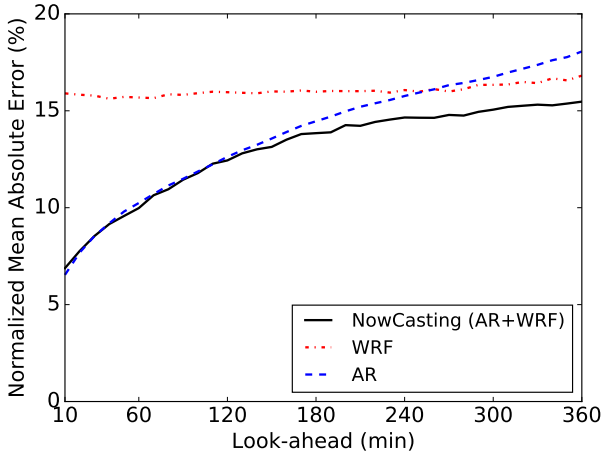
For each of the three prediction models, Fig. 2.7(a) shows the NMAE as a function of look-ahead time, which is the prediction horizon expressed in minutes. The horizontal axis



(a) Comparison of NowCasting results from 3 models



(b) Model A_PD_S



(c) Model A_SD_S

Figure 2.7: Prediction results for the average turbine for different look-ahead times.

is the rolling prediction window with look-ahead times of 10 minutes up to 6 hours (360 minutes). The NMAE is on the vertical axis and naturally increases with the look-ahead time. That effect is less apparent in the WRF predictions because the numerical weather prediction model is based not on real-time measurements taken at the turbine location, but on aggregated weather data at a coarser granularity.

The NowCasting approach is a weighted combination of the autoregressive (AR) and Weather Research and Forecasting (WRF) approaches. It essentially fuses those two approaches in a manner that improves the prediction performance, as seen in Figs. 2.7(b) and (c). The weights vary for each look-ahead time, and are estimated using neural networks or

the L^1 linear regression model.

We evaluated different neural network configurations by increasing the depth from zero (linear) to 6 densely connected hidden layers. The $A_P D_S$ model was found to benefit from a nonlinear mapping because the WRF input is a speed measure while the output is a power measure. When the inputs and outputs were both in the power domain (as in $A_P D_P$), we found that deeper neural networks performed worse than the linear model. That also held true for $A_S D_S$, in which case the inputs and outputs were in the speed domain before conversion to the power domain. Therefore, for $A_P D_P$ and $A_S D_S$, the L^1 minimization approach (as discussed in Section 2.3.1) would have done as well as a neural network.

2.5.5 Impact of Feedforward Neural Network Depth

Illustrations of the depth comparisons are in Fig. 2.8. Nonlinear models work best for $A_P D_S$, but linear models work best for $A_P D_P$ and $A_S D_S$.

2.5.6 Investigating the Impact of Time and Date

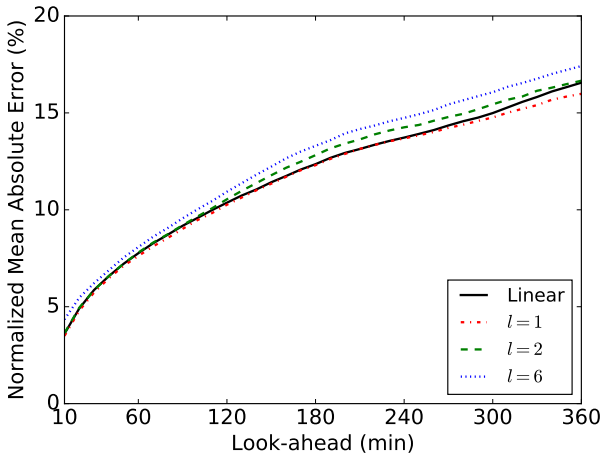
Thus far, we have illustrated results in which averaging removed the influence of time and date. In this subsection, we take a closer look at the impact of time and date on the results.

Impact of Time of Day

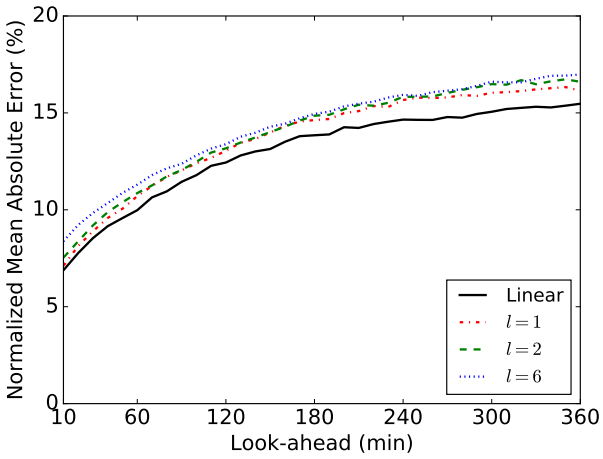
Using the notation in Section 2.5.2, we obtain the NMAE for the time of day as follows.

$$\text{NMAE(k)} = \frac{1}{|\tau||\eta|} \sum_{i \in \eta, j \in \tau} \frac{|f_p^*(X_{i,j,k}) - y_{i,j,k}|}{r}, k \in \rho. \tag{2.5}$$

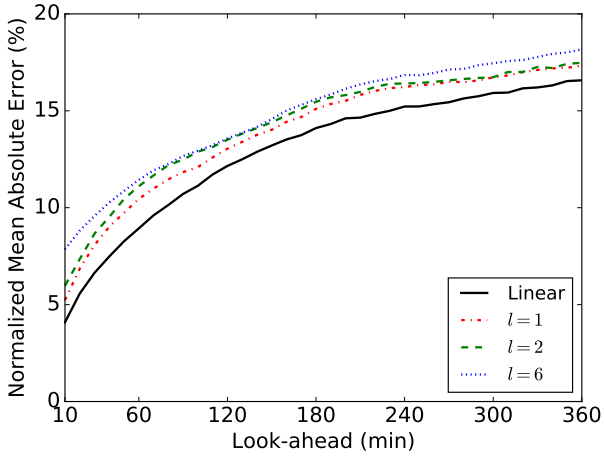
The results are illustrated in Fig. 2.9 for the $A_P D_S$ model. We notice that the prediction suffers the most between 9:00 P.M. and 1:00 A.M., when the wind speed is particularly erratic.



(a) Model $A_P D_S$



(b) Model $A_S D_S$ + Power Curve



(c) Model $A_P D_P$

Figure 2.8: Neural network depth comparisons for different models on the mean turbine.

Impact of Day of Year

Using the notation in Section 2.5.2, we obtain the NMAE for the day of the year as follows.

$$\text{NMAE(i)} = \frac{1}{|\tau||\rho|} \sum_{j\in\tau, k\in\rho} \frac{|f_p^*(X_{i,j,k}) - y_{i,j,k}|}{r}, i \in \eta. \tag{2.6}$$

Since we chose cross-validation groups spread evenly across the year, we were able to investigate the impact of the day of the year. The results are illustrated in Fig. 2.10 for the $A_P D_S$ model. We notice that the prediction suffers the most for May 2015, October 2015, and February 2016. Upon investigating the possible causes of those anomalies, we found that

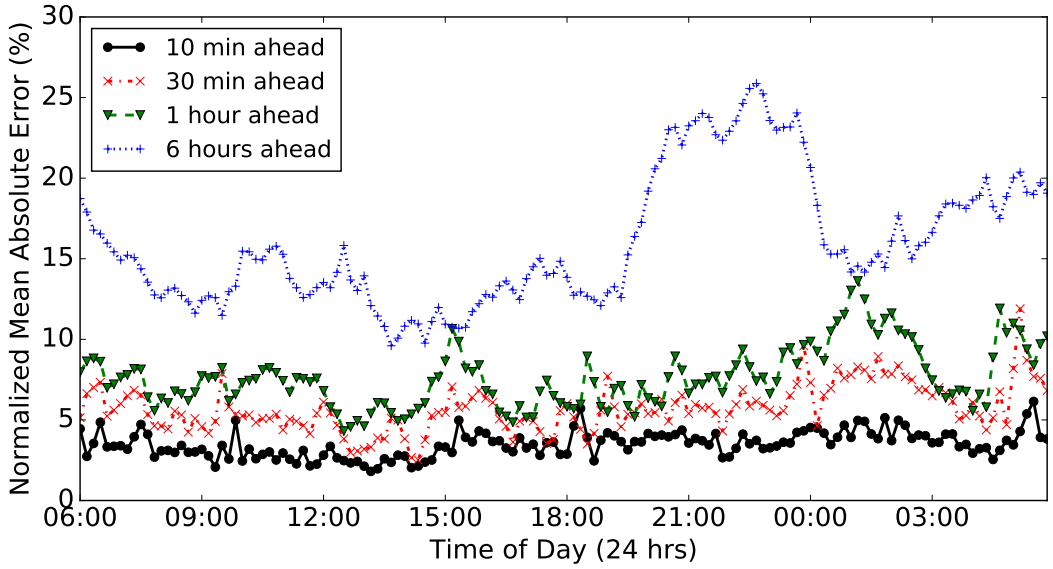


Figure 2.9: Impact of time of day on prediction results.

storms were reported during those three months (see [35], [36], and [37]). As our training set did not prepare the model for the strong winds and gusts seen in the test set during those months, the prediction accuracy suffered.

2.5.7 Investigating the Results across Turbines

In Fig. 2.7, we illustrated the prediction accuracies for the "average" turbine. In Fig. 2.11, we illustrate the minimum and maximum errors reported by individual turbines averaged across all cross-validation groups. It can be observed that the range of errors increases with the look-ahead period, as expected because of the increased uncertainty.

2.5.8 Results for Wind Farms on Aggregate

We obtained the prediction accuracy for the wind farm as a whole, aggregating the 21 wind turbines as follows.

$$\text{NMAE of Wind Farm} = \frac{1}{r|\tau|}.\frac{1}{|\rho|}\sum_{k\in\rho}|\sum_{j\in\tau}[f_p^*(X_{i,j,k}) - y_{i,j,k}]|, \tag{2.7}$$

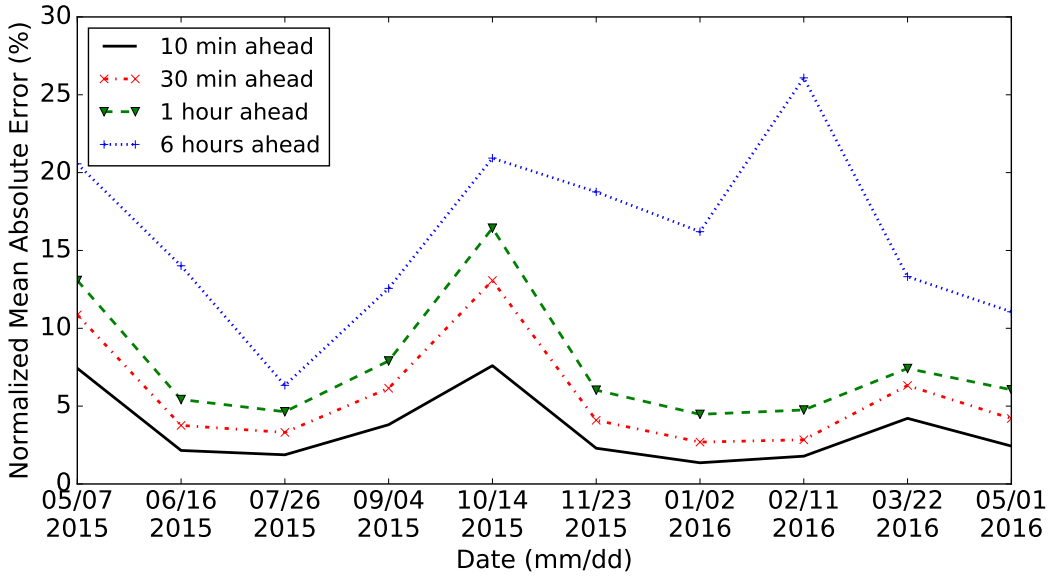


Figure 2.10: Impact of day of year on prediction results.

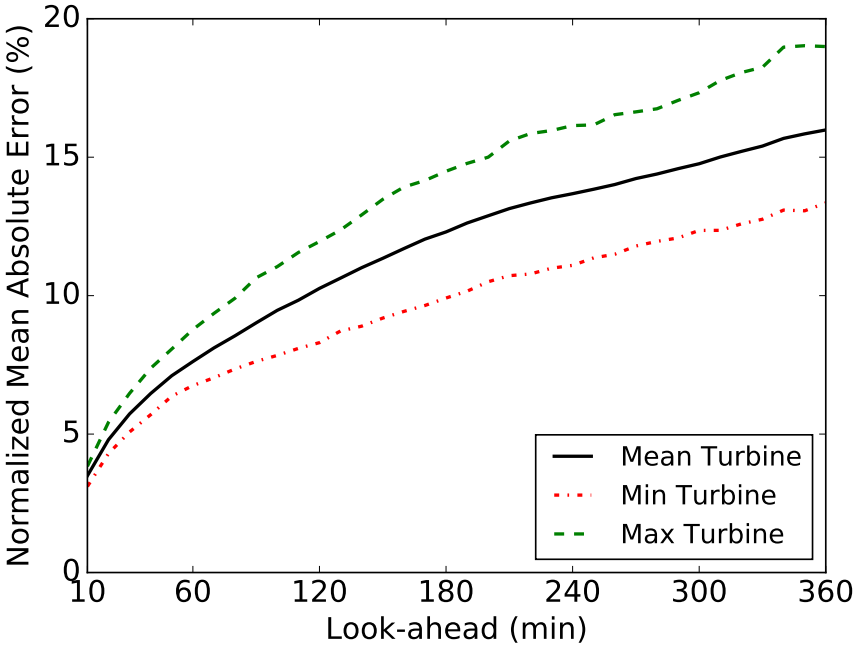


Figure 2.11: Spread of prediction results across turbines.

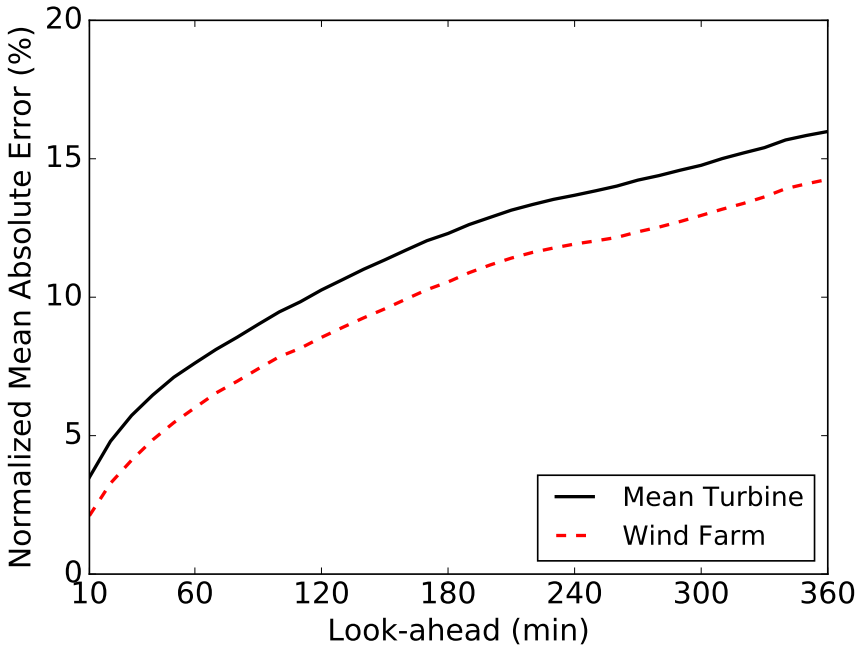


Figure 2.12: Prediction results for wind farm on aggregate.

where normalization was done by $r|\tau|$, which is the total capacity of the wind farm. The final result was averaged across all cross-validation groups in η.

Notice that Eq. (2.7) is different from Eq. (2.4). In Eq. (2.7), the errors are added before the absolute value is taken. The results are illustrated in Fig. 2.12. As expected, the errors for the wind farm as a whole were lower than the errors for the mean turbine. The reason is that positive errors from one turbine canceled out negative errors from a different turbine in the same farm.

2.5.9 Impact of Size of Training Set

We trained, validated, and tested our model on 18, 2, and 1 day(s) of data, respectively. Practitioners may be interested in knowing how much data they need to collect in order to perform predictions with high accuracy, so we present an evaluation of the prediction accuracy for training sets of different sizes in this section.

We evaluated 10, 20, and 40 days of data to fit our model. As we kept the validation split of 10% (described in Section 2.5.2), those days corresponded to training set samples

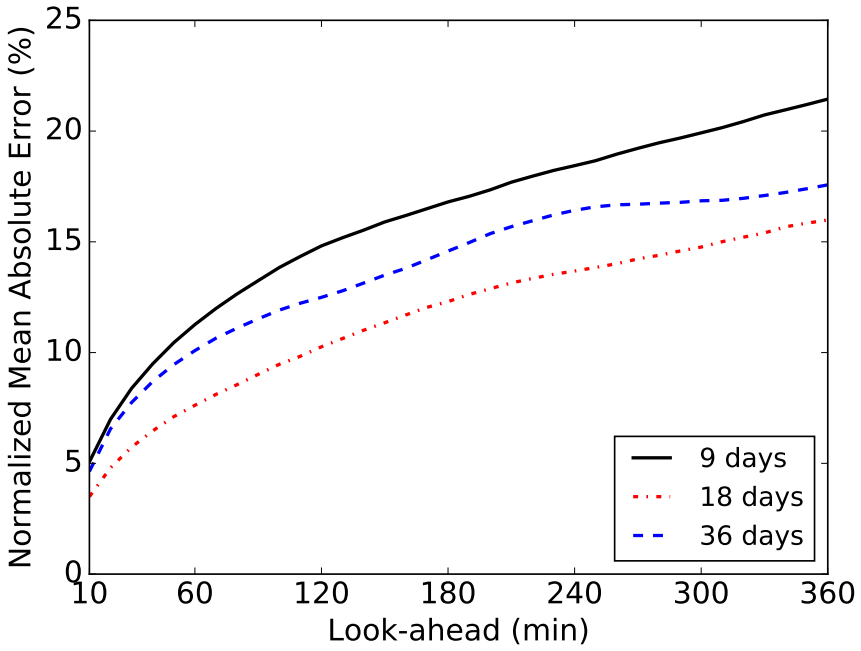
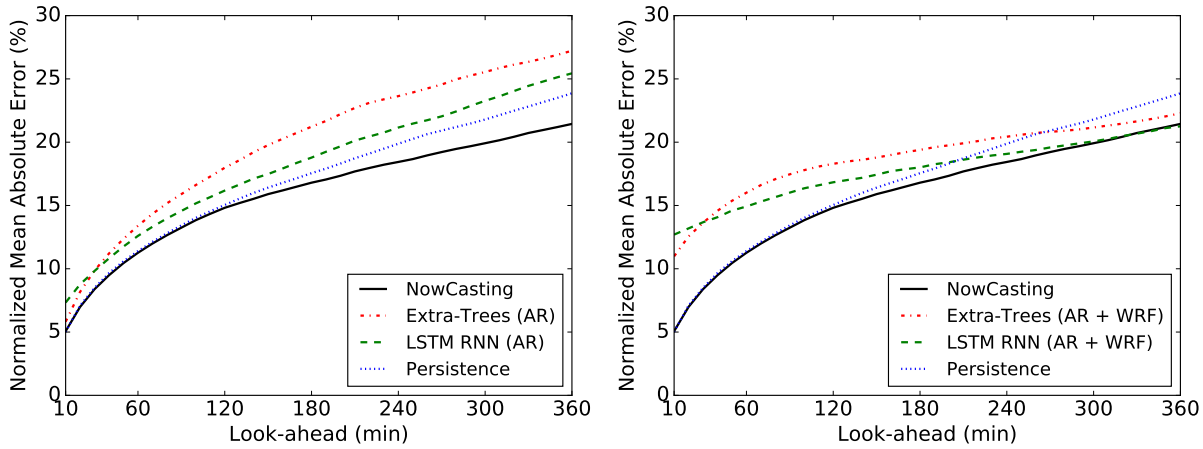


Figure 2.13: Impact of training set size on prediction accuracy.

of lengths 9, 18, and 36 days, respectively. The remaining 1, 2, and 4 days were used for validation. We maintained the test set size at 1 day in all scenarios. As illustrated in Fig. 2.13, having 18 days for training produced the best results. Nine days of data were not enough to accurately fit the model. On the other hand, 36 days of data captured obsolete temporal patterns that were not useful in making predictions for the immediate future. As a result, the prediction accuracy suffered.

2.5.10 Comparisons with Related Work

In this section, we compare our methods with the purely auto-regressive approaches proposed in related work. In general, our approach does not significantly outperform the autoregressive approach for look-ahead times of 10–30 min, and is comparable to approaches used by the authors of [26], [27], and [38]. For longer look-ahead times (beyond 3 hours), our approach provides significant improvements over those approaches. Note that our approach combines the additional data from WRF using feedforward neural networks. In Fig. 2.14(a), we show the advantage of using feedforward neural networks and the additional WRF data.



(a) NowCasting (AR + WRF) vs. related work (AR)

(b) NowCasting (AR + WRF) vs. related work (AR+WRF)

Figure 2.14: Comparisons between alternative machine learning algorithms and feedforward neural networks. The alternative algorithms use only WRF readings in (a) and use WRF and AR readings in (b). The persistence model uses only AR readings.

In Fig. 2.14(b), we show that even if alternative machine learning techniques were to be trained using the additional WRF data, they do not perform as well as feedforward neural networks do. With the additional WRF data, the alternative machine learning techniques achieve improved accuracy looking 6 hours ahead, but suffer loss of accuracy in shorter time horizons. Feedforward neural networks combine the WRF and AR data effectively for all time horizons.

Some of the approaches proposed in related work are so memory-intensive that they could not be trained on 18 days of data on our server with 32 GB of memory. Therefore, we used 9 days of training data, 1 day of validation data, and 1 day of test data to make the comparisons among the approaches in this subsection.

Persistence Model

The persistence model is commonly used as a comparison baseline in various approaches in the literature, including those surveyed in [39]. In that model, the predicted value for the next time period is assumed to be equal to the value in the current time period. The results are illustrated in Fig. 2.14(a). Although the prediction accuracy gains of NowCasting

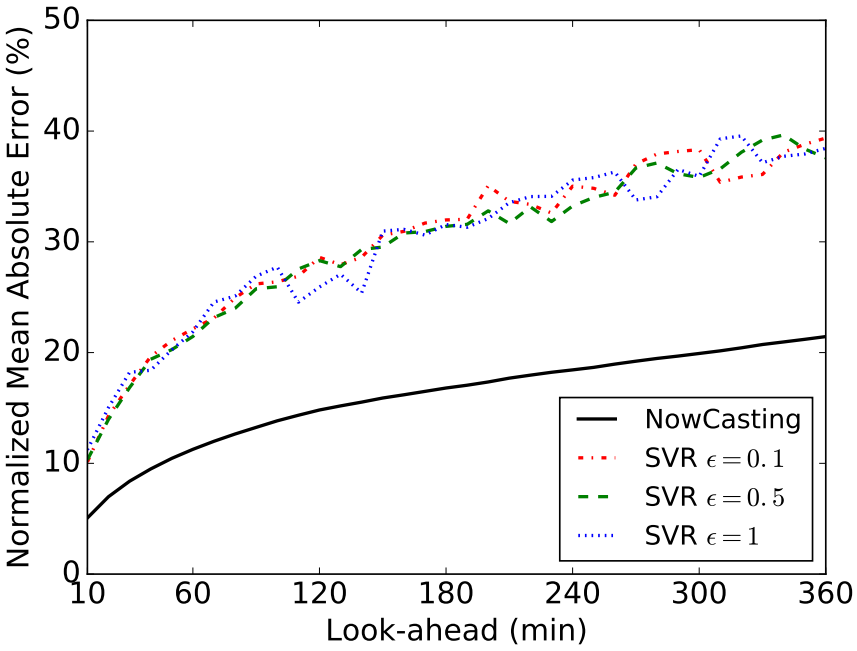


Figure 2.15: NowCasting vs. Support Vector Regression.

over the persistence model are modest, NowCasting dramatically improves the prediction precision, and we will discuss that later, in Section 2.7.

Extremely Randomized Trees

Random forests were recently used to perform hour-ahead predictions in [38]. Random forests are obtained by averaging decision-trees trained on multiple subsamples of the training data. Extremely randomized trees (extra-trees) incorporate additional randomization in the learning algorithm of each decision tree. They have a major disadvantage with respect to neural networks in that they consume an immense amount of memory. We were unable to fit an extra-trees regressor with more than five trees even on our reduced training set of 9 days of data. The reason is that the algorithm, which was provided in Scikit-Learn [40] and configured to minimize the MAE, was not able to converge with our 32 GB of memory for larger forests.

Recurrent Neural Networks

Long short-term memory (LSTM) recurrent neural networks (RNNs) were first devised by the authors of [41]. They were used in [18] and [42] for wind power prediction. We found that LSTMs, like extra-trees, had a huge memory requirement, occupying a full 32 GB of memory for training, whereas feedforward neural networks occupied only 400 MB. LSTM layers have many more parameters to fit than do normal, densely connected layers. We believe that that led to overfitting, leading to relatively inaccurate results, as illustrated in Fig. 2.14.

Support Vector Regression

The authors of [26] used the Support Vector Regression (SVR) method for forecasting wind power. SVR has both linear and nonlinear approaches. The nonlinear approaches provided in LibSVM [43] have well-known scalability issues, as stated in [44], and did not converge even on our reduced training set that contained only 9 days of data. Therefore, we compared NowCasting with linear SVR, which is scalable.

In Fig. 2.15, we illustrated the results of SVR for different values of the loss parameter, ϵ, which is described in [44]. SVR, by design, predicts each horizon independently of other horizons. The reason is that the ground truth must have a dimensionality of 1. For all other machine learning methods, the ground truth can have a dimensionality of h, so all horizons are simultaneously (not independently) considered in the training. That is the reason for the large fluctuations seen in the curves for SVR, which are absent in the other approaches. SVR was found to perform the worst of all the methods, with and without the additional data from the WRF.

ARMA Model

The forecasting approach in [45] uses the ARMA model. We used the R forecasting libraries [46] to perform ARMA forecasting, but found through those libraries that our dataset was not (wide-sense) stationary. In other words, the mean, variance, and autocovariance

changed over time. As a result, the ARMA model would not be a suitable model for wind data, although it is well-suited for predicting consumer demand, as was done using the same R libraries in [47]. We have already made comparisons with AR models (ARMA without the moving average component) in Fig. 2.7, so we do not illustrate them again in this section.

2.6 Verification of Models

In this section, we show that our evaluation results are theoretically consistent. In minimizing the MAE we inherently assume that the errors are distributed as a Laplace distribution. We now show that that is the case. The general formulation Eq. (2.1) for MAE minimization can be rewritten as a maximization problem, given as follows.

$$f_P^* = \arg\max_{f_P} \frac{1}{m} \sum_{i=1}^{m} -\frac{|f_P(X_i) - y_i|}{b}, \tag{2.8}$$

where $b > 0$ is some positive constant. Since $e^{mx}/(2b)^m$ strictly increases in x, equation Eq. (2.8) can be written as

$$f_P^* = \arg\max_{f_P} \prod_{i=1}^{m} \frac{1}{2b} \exp\left(-\frac{|f_P(X_i) - y_i|}{b}\right).$$

Therefore, finding f_P^*, which minimizes the MAE, is equivalent to finding a function f_P that maximizes the likelihood that the prediction error $\epsilon_i = f_P(X_i) - y_i$ is distributed as a Laplace distribution with mean zero and variance $2b^2$. The variance can be used to quantify the uncertainty associated with each prediction, which we discuss in Section 2.7.

Therefore, it is sufficient to verify that the prediction errors that we observe from using our prediction models are indeed distributed as Laplacian. If they are not, we would know that our neural network model was not suited to minimizing the MAE. We chose the P-P plot approach to show in Fig. 2.16 how close the error distribution is to the theoretical Laplace distribution, for one turbine and 1-hour-ahead predictions. The P-P plot is a scatter plot of the empirical CDF of the provided data as a function of the theoretical CDF of the Laplace distribution. It can be seen in Fig. 2.16 that the Laplacian distribution is a good

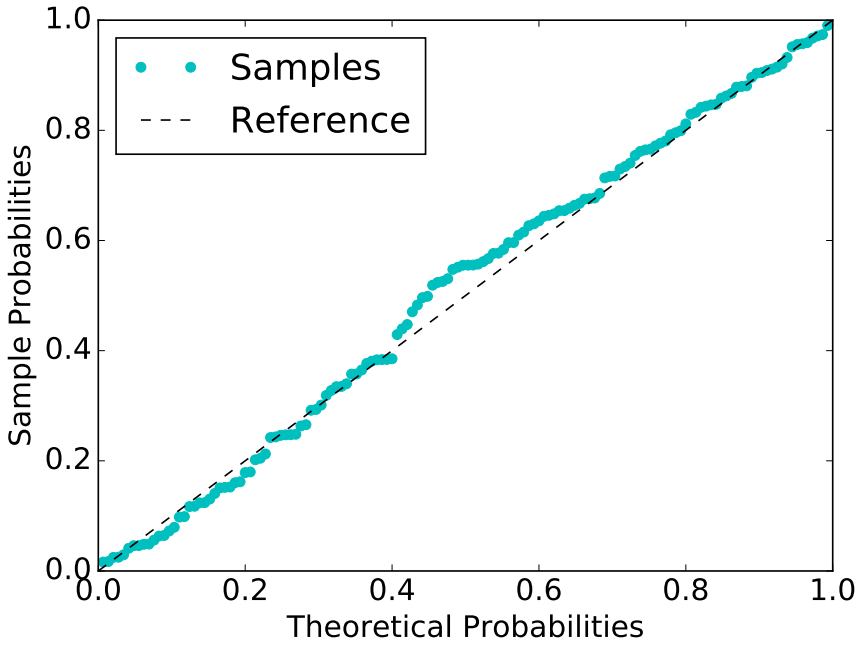


Figure 2.16: P-P plot comparing prediction error distribution for a single turbine, 1 hour ahead, against the theoretical Laplace distribution.

approximation for the empirical error distribution, verifying our neural network model.

2.7 Quantifying Prediction Uncertainty

In addition to improving prediction accuracy, we seek to increase the confidence in our predictions by improving precision. While accuracy is concerned with keeping the mean of the prediction errors close to zero, precision is concerned with keeping the standard deviation of the errors small. Quantifying the precision allows utilities to quantify the risk associated with committing to a certain level of wind generation. For example, knowing that the prediction error is going to be between, say, -1 MW and 1 MW with 90% confidence will improve a utility's operational decisions, such as the scheduling of generators or dynamic loads. It can also help the wind farm operator make a more informed bid in the electricity market, reducing the risk of curtailment costs.

The confidence intervals for error estimates can be obtained empirically using percentile points on the observed distribution of prediction errors. Alternatively, a model-based ap-

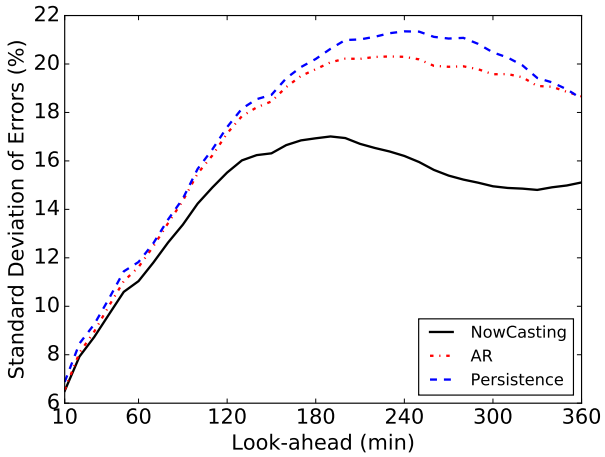
proach can be used by taking percentile points of the Laplace distribution, given that the prediction errors obey that distribution. We use σ to denote the standard deviation of the errors, and $\sigma = \sqrt{2}b$ for the Laplace distribution. $F_\epsilon(\alpha) = P(\epsilon < \alpha)$ denotes the CDF of the continuous distribution. As our errors have mean zero, $F_\epsilon(\alpha) = \frac{1}{2}e^{\alpha/b}$, for $\alpha < 0$, by definition of the Laplace distribution. For $\alpha_p > 0$, the confidence interval for the symmetric Laplacian can be obtained as follows.

$$P(-\alpha_p < \epsilon < \alpha_p) = 1 - 2F_\epsilon(-\alpha_p) = 1 - e^{-\sqrt{2}\alpha_p/\sigma} = p, \tag{2.9}$$

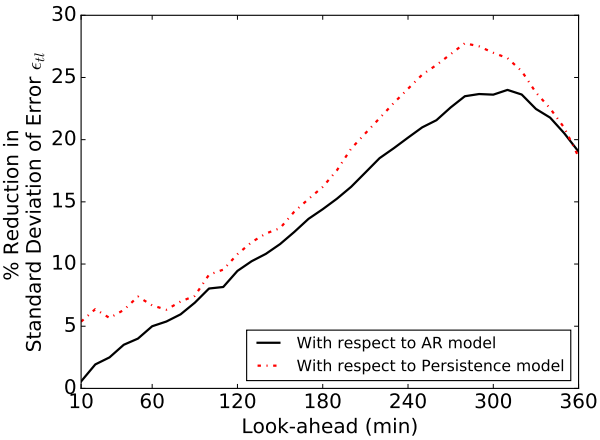
where p is the confidence level associated with an interval of width $2\alpha_p$. For example, $p = 0.95$ would produce a 95% confidence interval of width $2\alpha_{0.95}$. If we were to reduce σ by a factor of k, then we would be able to decrease the width of the confidence interval by a factor of k and obtain the same confidence level p.

The standard deviations of errors in predicting wind farm generation (σ) are illustrated for the autoregressive approach, persistence model, and NowCasting ($A_P D_S$) approach in Fig. 2.17(a). The standard deviations are in megawatts, but we normalized them with respect to the farm capacity and expressed them as a percentage for improved readability. The standard deviations increase with increase in look-ahead, and then converge because the errors are bounded by the wind farm's rated capacity.

NowCasting has a clear advantage over the AR and persistence models, as illustrated in Fig. 2.17(b). It reduces uncertainty by decreasing the width of the confidence intervals by over 20% for predictions made 4 hours into the future. Conversely, prediction errors made with NowCasting are more likely to lie within a fixed confidence interval than are errors made with other approaches. In that sense, *NowCasting is a more reliable prediction approach* that can help utilities and wind farm owners reduce the risk and cost associated with scheduling or bidding for a specific amount of wind power generation.



(a) Normalized standard deviation



(b) % Reduction achieved by NowCasting

Figure 2.17: Standard deviations of prediction errors are directly proportional to the uncertainty (width of confidence intervals).

2.8 Improving Utilization of Wind Power

Having shown that, by using NowCasting, the width of the confidence intervals of the wind power predictions can be reduced by 20%, we now illustrate how that can help improve the utilization of wind power. In an electricity market, the wind farm operator must place a bid to provide a specified amount of wind power at the market price. If the wind farm operator were to under-deliver, they would be subjected to heavy penalties. If the wind farm were to over-deliver, the excess power would be curtailed and incur operational costs. Therefore, the operator's objective is to place the highest reliable bid possible.

Consider a scenario in which the wind farm operator wants to place a bid that can be satisfied with a 99% reliability. We use B to denote the bid amount and G to denote the actual generation. Then, we need to find B such that

$$B = \arg\max_{B'} P(G < B') \leq 0.01. \tag{2.10}$$

We construct a symmetric 98% confidence interval on the prediction errors, such that 1% of the error probability remains in the lower and upper tails of the Laplacian distribution of errors. We use ϵ_N to denote the prediction error obtained through the use of NowCasting, and ϵ_P to denote the corresponding error from the use of the persistence model. We use

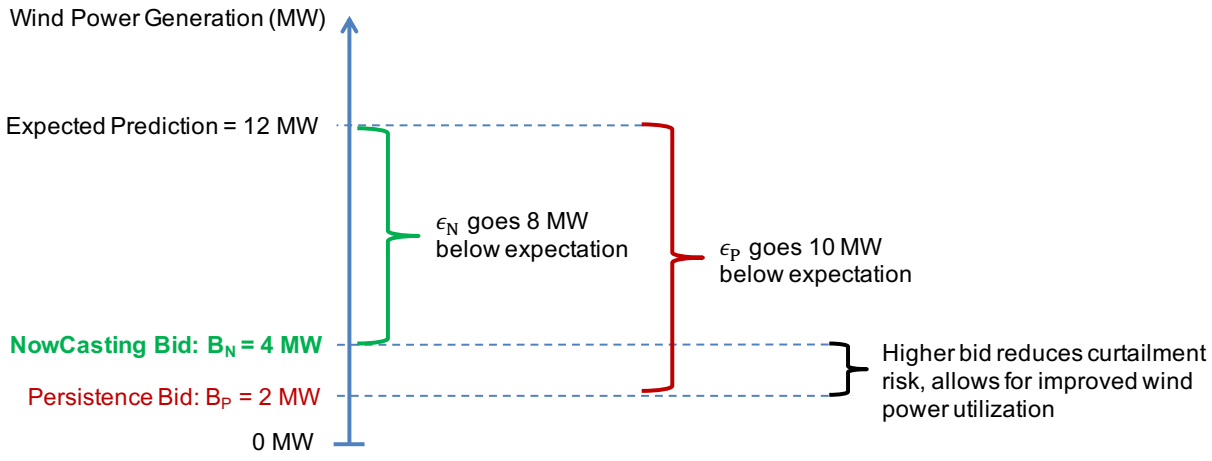


Figure 2.18: NowCasting can be used to place a higher bid than the persistence model, allowing for better wind power utilization.

B_N and B_P to denote the bids that would be placed using the NowCasting and persistence models, respectively.

So far, our notation and requirements have been general. For illustration, we will take a concrete example wherein the expectation of predicted wind power, by using both NowCasting and the persistence model, is 12 MW. Assume that the 98% confidence interval of ϵ_P is given by the range $[-10 \text{ MW}, 10 \text{ MW}]$. NowCasting reduces that width by 20% to the range $[-8 \text{ MW}, 8 \text{ MW}]$ for ϵ_N. As illustrated in Fig. 2.18, the solution to Eq. (2.10) is given by $B_P = 2$ MW and $B_N = 4$ MW. While B_N would always be greater than B_P, in that example, it was twice as large. Thus, the operator would be able to place a higher bid using NowCasting and get the same level of reliability as the they would have through the use of the persistence model. That reduces the curtailment risk and allows the operator to be paid more for the additional 2 MW supplied.

2.9 Related Work

In this section, we discuss the literature on short-term wind power prediction. A more detailed survey is presented by Foley et al. [39]. Our work is different from most related work in one important aspect: we make predictions at a time granularity of 10 minutes, while most related approaches make predictions at a granularity of an hour. It is more difficult to

make predictions for smaller time granularities because errors average out over larger time periods.

Several authors have used various neural network configurations to estimate prediction functions from large data sets, as surveyed in [39]. The most common configuration is the feedforward neural network in a NAR model with one hidden layer, commonly referred to as a *3-layer perceptron* [26,27]. The authors of [27] present results for very short-term forecasts (up to 30 minutes ahead). None of the prior efforts that used the 3-layer perceptron [26,27] presented a detailed study of the consequences of varying the depth of the neural networks. We present such a detailed study and find that a single layer is sufficient.

Kusiak et al. [26] evaluate wind farm-level forecasting methods (looking 10 minutes to 4 hours ahead) using machine learning tools, such as the Boosting Tree algorithm for feature selection, Support Vector Regression, the Bagging Tree, the M5P Tree, the Reduced Error Pruning Tree and the 3-layer perceptron. The authors claim that their approach for predicting wind speed yields poor results, whereas our approaches yield highly accurate results for predicting wind speed that translate well into wind power prediction. Random forests were used to perform hour-ahead predictions in [38].

Most of the literature [18,26,27,38,48,49] on predicting wind power assumes AR or NAR models, instead of also incorporating exogenous inputs. NARX models are used in [19] for predicting wind power days ahead, where the exogenous input is temperature. Cadenas et al. use a NARX model with exogenous inputs from solar radiation measurements to improve the precision of 1-hour-ahead predictions [50]. Instead, we take exogenous wind speed inputs from a hyperlocal weather forecasting engine.

Men et al. also use neural networks that combine exogenous inputs from a WRF engine that has a spatial resolution of 1.8 miles [51]. Their predictions are for a 72-hour-ahead period, and at a time granularity of 1 hour (not as fine-grained as ours). Their evaluation does not consider the decrease in accuracy as the prediction horizon increases. Also, they do not perform cross-validation, and therefore their results may be biased towards specific times of the year.

Another drawback of the approaches in related work pertains to how the neural network models were trained. The authors of [27] trained their networks using least-squares mini-

mization or Bayesian learning [27], as opposed to L^1 error minimization, when the required prediction metric to be minimized was the L^1 error. In this chapter we take a more rigorous approach by recognizing that if a particular metric is used to test a prediction method, that same metric must be used to train it. Otherwise, the results would be suboptimal.

Barbounis et al. use local recurrent neural networks represented as feedforward neural networks to predict long-term [18] and short-term [42] wind power. The authors of [45] use the ARMA model to forecast wind power. Our approach does not make the assumption implicit in ARMA models that the data are weakly (or wide-sense) stationary. In fact, we were unable to fit our dataset to an ARMA model because our data were not wide-sense stationary.

The application of machine learning with historical and weather data to short-term solar predictions has been explored in the literature [52–55]. Our paper and [26, 27, 56] explore similar ideas for short-term wind predictions. Unlike wind power, solar power generation exhibits a strong diurnal pattern. Therefore, the approaches used to forecast solar power may not be directly transferable to wind power.

2.10 Conclusions

In this chapter, we showed that empirical models of wind power generation, constructed using machine learning methods, improved the utilization of wind power as an energy resource. Our approach improves on not only the baseline heuristic model (persistence model), but also other machine learning approaches proposed in the literature.

The proposed approach improved both accuracy and precision of short-term wind power predictions. It used ARX/NARX models with exogenous inputs from a hyperlocal forecasting engine to achieve that improved accuracy. We evaluated the performance of the proposed approach and compared it to existing approaches. The proposed approach achieves a normalized mean absolute error of 2.11% when looking 10 minutes ahead, and 14.25% when looking 6 hours ahead, for a wind farm of 21 turbines. The key benefit of our approach is that it strikes the right balance between the autoregressive model, which is accurate for forecasts less than 1 hour ahead, and a WRF model, which is accurate for forecasts greater

than 5 hours ahead.

Our approach not only improves the prediction accuracy, but also decreases the uncertainty associated with the predictions by over 20% of that of approaches in related work. That allows utilities to evaluate the risk associated with scheduling wind power. Finally, we presented the most extensive evaluation of wind power prediction that we know of, while providing researchers and practitioners alike with key insights on how to develop, implement, and verify working solutions. The work in this chapter was peer-reviewed and published in [57].

CHAPTER 3

DATA-DRIVEN DEMAND RESPONSE

In this chapter, we continue the theme from Chapter 2 of using data to improve resource utilization in power grids. In particular, we present a data-driven approach to quantify energy waste in buildings and the ability of buildings to provide demand response (DR). This chapter contains work that was done by the author as part of a team at the Advanced Digital Sciences Center (ADSC). The chapter focuses on the author's contributions to the body of work; the full body of work was published in [58] and [59]. Key insights provided by Dr. Deokwoo Jung, who led the research at ADSC, have been included (and explicitly attributed) to provide context to the author's contributions.

DR is the ability of buildings to modify their electricity demand in response to operator signals. The buildings may be residential, commercial, or industrial, and the signals may come from regional grid operators or from electric utilities. The need for demand response has arisen mainly because of the inability of the electric power grid to adapt to the fast-growing demand. As a result, there is either insufficient generation capacity, or insufficient ability of transmission/distribution lines to transport the power necessary to meet high demand. High demand is usually a problem only during peak hours of the day when most consumers simultaneously need power. It is not a problem during the night hours, when consumers tend to use less power. As a result, assets such as transmission/distribution lines need to be overprovisioned to support peak load, and are underutilized in off-peak periods.

3.1 Summary of Contributions

In this chapter we make two main contributions. First, we quantify the extent to which buildings can reduce their electricity consumption while ensuring occupant comfort. Second,

we use that ability of buildings to reduce electricity consumption to model the capacity of buildings to provide DR by reducing load by specific amounts.

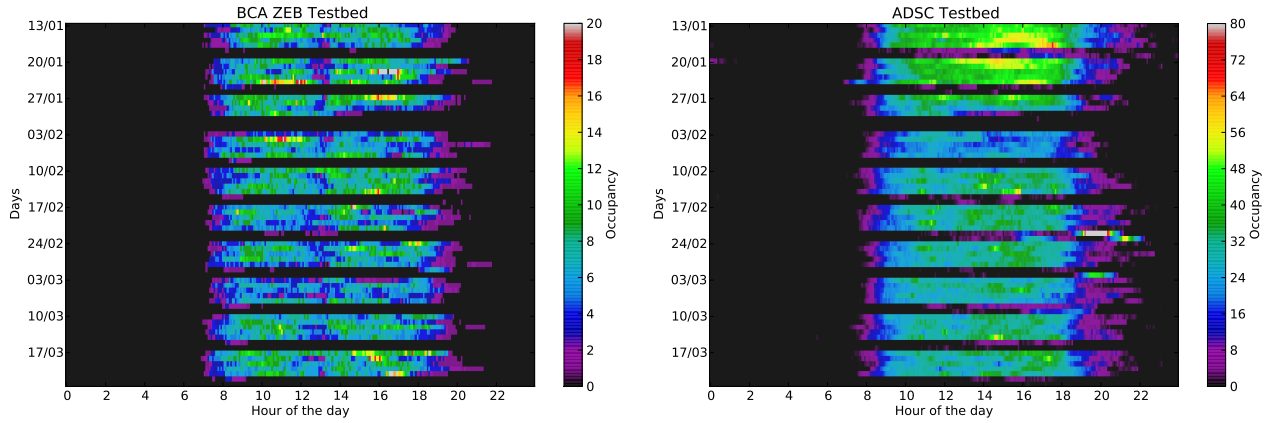
We developed a sensor-driven energy use analysis system called *EnergyTrack*, which continuously analyzes, evaluates, and interprets energy usage in buildings using real-time sensor data. The system uses an analytical energy usage model that considers the energy consumption of individual building loads, occupancy of building zones, and occupant comfort level. This model has two major advantages over existing models, such as [60], which only consider static baseline consumption to quantify energy savings. First, our model naturally accounts for the trade-off between energy savings and their impact on occupant comfort. Second, it also favors energy utilized during periods of high occupancy over periods of low occupancy. These two features quantitatively incorporate the intuition that energy usage efficiency of a load is high when occupancy and occupant comfort are high.

An efficient DR program should take full advantage of the DR potential of each participating consumer. In order to facilitate such DR, it is crucial to determine the maximum reduction in energy consumption that can be reliably achieved by buildings. In our model, that reduction in consumption can only be achieved as long as the comfort criteria of building occupants can be met despite that reduction. We refer to the reduction in consumption as the *demand response capacity* of the building, and it is stochastic because it is a function of stochastic parameters such as occupancy and weather. Our model provides a measure of reliability associated with the DR capacity, and that is based on a probabilistic model of energy consumption with respect to occupancy and weather.

3.2 Description of Testbeds

For this work, we deployed wireless sensor networks (WSNs) inside two office spaces in Singapore to gather data on occupancy, occupant comfort, and energy usage. These data were used to develop a DR capacity model that incorporates occupant comfort constraints. In this section we provide details about the two testbed sites and about our sensor network deployment.

The testbeds are office spaces in two different commercial buildings. **Testbed #1 (ZEB**



(a) ZEB testbed: estimated occupancy

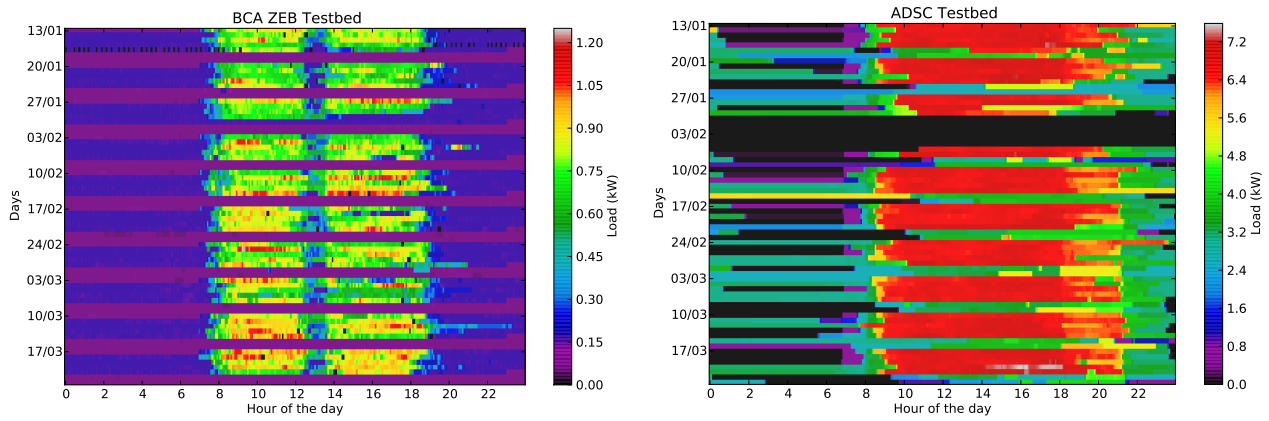
(b) ADSC testbed: estimated occupancy

Figure 3.1: Real-time occupancy in two office testbeds over a 10-week period in 2014.

testbed) is an office space in a net-zero energy building (ZEB) [61] owned by the Building and Construction Authority (BCA) of Singapore. The testbed is located in the office of BCA's Centre for Sustainable Buildings and Construction. The ZEB has several energy-saving features that are not common in most buildings today, such as light pipes and light shelves (to bring more daylight into the indoor space), as well as displacement ventilation and personalized fresh air supplies at individual desks. **Testbed #2 (ADSC testbed)** is the office at ADSC. This testbed is typical of a modern commercial office building in any modern city that has warm weather throughout the year. The exterior wall is made of double-layered glass.

We instrumented both testbeds with plug meters, temperature-humidity-light (THL) sensors, CO_2 sensors, and passive infrared (PIR) sensors. The WSNs consisted of motes running TinyOS with a TI MSP430F1611 MCU (8 MHz clock rate and 10 KB RAM) and a Chipcon CC2420 Zigbee radio.

The ADSC testbed was additionally instrumented with power meters in the main switch board (MSB). Those meters monitored 42 electrical branches that were grouped into plug-load, lighting, and server-room use. They allowed us to analyze consumption by load types. The plug meters were installed at individual desktop computers and reported power consumption every second by default.



(a) ZEB testbed: power consumption from lighting (b) ADSC testbed: power consumption from lighting

Figure 3.2: Measured lighting power consumption from two office testbeds over a 10-week period in 2014.

3.2.1 Empirical Comparison of Testbeds

In addition to differences in construction, the two testbeds differed markedly in size, operating hours, and occupancy patterns (see Fig. 3.1). The ZEB testbed occupied a 154.5m^2 area on a single floor of the building, and typically housed 10 to 12 people working during the day. In contrast, the ADSC testbed was much larger (although still on a single floor), at 824.5m^2, and had a typical occupancy of roughly 40 people during working hours. The working hours at the ADSC testbed were less regular than in the ZEB testbed, with the space often occupied past 8:00 P.M. In both spaces, personal computers were the prevalent plug loads. ZEB employees used laptops, whereas ADSC employees used desktops (desktops use more energy than laptops).

In the ZEB testbed, lighting power consumption data were obtained from the building management system (BMS). In the ADSC testbed, the BMS data were not made available, so lighting power was measured from the electrical distribution panels. Figure 3.2 shows heat maps depicting the measured lighting power consumption in each testbed over the 10-week data collection period. Similarly, Fig. 3.3 shows the average lighting power per unit floor area for weekdays in each of the testbeds. The impact of daylight and dimmable lighting in the ZEB testbed is apparent from the significantly lower lighting power density.

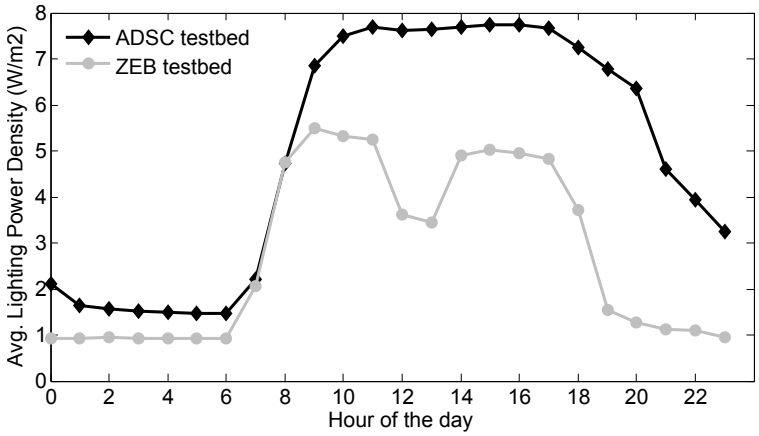


Figure 3.3: Average lighting power consumption per unit of floor area on weekdays, over a 10-week period in 2014.

3.2.2 Sensor Data Storage

From all the sensors combined, we collected and stored in a MySQL database approximately 6.7 million data points every day. Different database designs were tested, and designs favoring fast read times were ultimately chosen over those that increase modularity. We maintained a DB table for each sensor node to reduce the SQL query execution time. We also took averages of sensor data every 15 minutes and stored them in separate tables. This caching method greatly reduced the execution time for search queries with acceptable storage redundancy.

3.3 Models Derived from Sensor Data

In this section we present models derived from sensor data that were used to quantify energy wastage and demand response capacity in buildings.

3.3.1 HVAC Model

In the ZEB testbed, the heating, ventilation, and air conditioning (HVAC) consumption was obtained directly from the BMS. In the ADSC testbed, however, the consumption data of the HVAC system was not available, as the entire BMS was confidentially managed by a private

building management company. Therefore, we used data generated from EnergyPlus [62] simulations, a method that is described in great detail in [63]. To drive realistic simulations, we created models in EnergyPlus that used detailed building specifications such as thermal envelope parameters, floor plan measurements, and air handler unit (AHU) specifications obtained from the building management company.

Our simulations revealed that the cooling capacity of the ADSC HVAC system was oversized: about three times larger than it needed to be. Such oversizing is common for commercial buildings because their HVAC systems are designed to handle worst-case cooling loads. As a consequence of the oversizing, the HVAC system operated inefficiently and drew a large, constant power regardless of occupancy and weather variations.

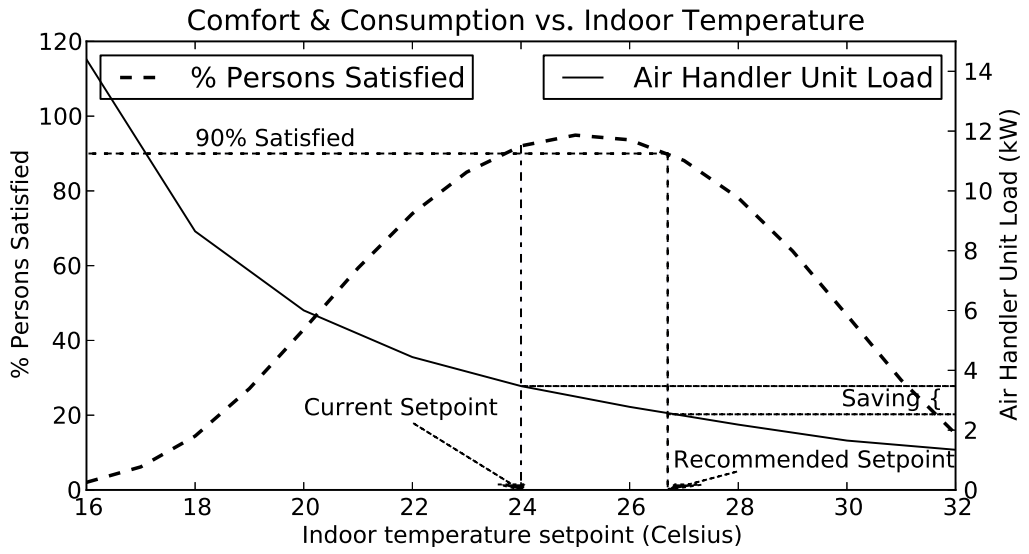


Figure 3.4: Plotting AHU power consumption as estimated using EnergyPlus, and thermal comfort as expressed as the percentage of persons satisfied with the temperature setpoint.

By running multiple EnergyPlus simulations to determine the AHU load for various indoor temperature setpoints, we obtained the AHU load curve shown in Fig. 3.4.

3.3.2 Comfort Model

We obtained the comfort associated with temperature settings, or *thermal comfort*, using the predicted percentage of dissatisfied people (PPD) standard [64]. The PPD metric is a

function that predicts the percentage of occupants who are dissatisfied with their thermal comfort, given various environmental and human conditions, such as air temperature, humidity, metabolic rate, and clothing insulation. Our system updates the PPD value hourly based on temperature and humidity data obtained from sensors. The PPD input parameters can be set through a user interface. We set them according to our observations of the occupants in the two testbeds, and plotted the PPD parameters for the ADSC testbed in Fig. 3.4. Figure 3.4 therefore illustrates the trade-off between the comfort and cost of the HVAC energy usage.

We computed the comfort associated with lighting, or *visual comfort*, using a logistic function that maps lux values measured by sensors to a scale from 0 to 1. All lux values above and below user-specified thresholds map to visual comfort values of 1 and 0, respectively. Intermediate values for visual comfort were obtained using logistic regression.

3.3.3 Occupancy Model

We estimated occupancy in both testbeds by using an auto-regressive (AR) model that combined measurements taken from CO_2 and PIR sensors. Since the ground truth for the real-time occupancy is difficult to obtain, we did not use regression to obtain the AR coefficients. Instead, we assumed that the occupancy followed a truncated Gaussian distribution wherein the minimum is zero and the maximum is the capacity of the building. We estimated the AR coefficients using that distribution and the expectation maximization (EM) algorithm for linear regression with incomplete data [65]. We refer the interested reader to [58] for details of the model; the model is neither a contribution of the author, nor integral to this dissertation. The model was used to generate Fig. 3.1.

3.4 Quantifying Energy Wasted in Buildings

In this section, we present and evaluate an approach called *EnergyTrack* that quantifies the energy wasted in commercial buildings. EnergyTrack uses the HVAC, comfort, and occupancy models described in Section 3.3.

3.4.1 EnergyTrack Model

We describe the essential aspects of the EnergyTrack model, which were conceived by Dr. Deokwoo Jung, and refer the interested reader to [58] for additional details.

Let $e_n(t_1, t_2)$ denote the total energy consumption of load n in the time interval $[t_1, t_2]$. Then we say that

$$\begin{aligned} e_n(t_1,t_2) \;=\; & \underbrace{e_n^s(t_1,t_2)}_{\text{Static Consumption, } e_n^s} + \underbrace{\bar{q}_n(t_1,t_2)e_n^d(t_1,t_2)}_{\text{Useful Dynamic Consumption, } e_n^u} \\ & + \underbrace{(1-\bar{q}_n(t_1,t_2))e_n^d(t_1,t_2)}_{\text{Wasted Dynamic Consumption, } e_n^w}, \end{aligned} \tag{3.1}$$

where e_n^s denotes the static baseline consumption that cannot be controlled, and e_n^d denotes the dynamic consumption that can be controlled. The dynamic consumption is split into useful and wasted consumption by means of the mean usage factor of load n, $\bar{q}_n(t_1, t_2) \in [0, 1]$. $\bar{q}_n(t_1, t_2) = 1$ when load n is providing maximum comfort (or value) to the occupants who use it when they are present. In that case, no energy is wasted. If there are no occupants using the load, $\bar{q}_n(t_1, t_2) = 0$, and all the dynamic (or controllable) energy is wasted. $\bar{q}_n(t_1, t_2)$ is small when occupants are present but are not deriving comfort or value. That could happen, for example, when lighting is insufficient or the temperature of the room is too hot or too cold. The detailed formulation of $\bar{q}_n(t_1, t_2)$ is extraneous to this chapter, and can be found in [58].

Figure 3.5 illustrates the HVAC system wastage in the ADSC testbed. We calculated it using occupancy levels as well as thermal comfort parameters. Note that $\bar{q}_n(t_1, t_2)$ is set to 1 for all plug loads. For plug loads alone, e_n^s in Eq. (3.1) is nonzero, and is estimated by calculating the base load beyond which consumption cannot be reduced.

3.4.2 Baseline Heuristic Method

We evaluate the EnergyTrack model against a baseline heuristic method for the ADSC testbed. The baseline method corresponds to reducing the estimated wastage without the use of a control system. For lights, the wasted energy is given by the amount of energy

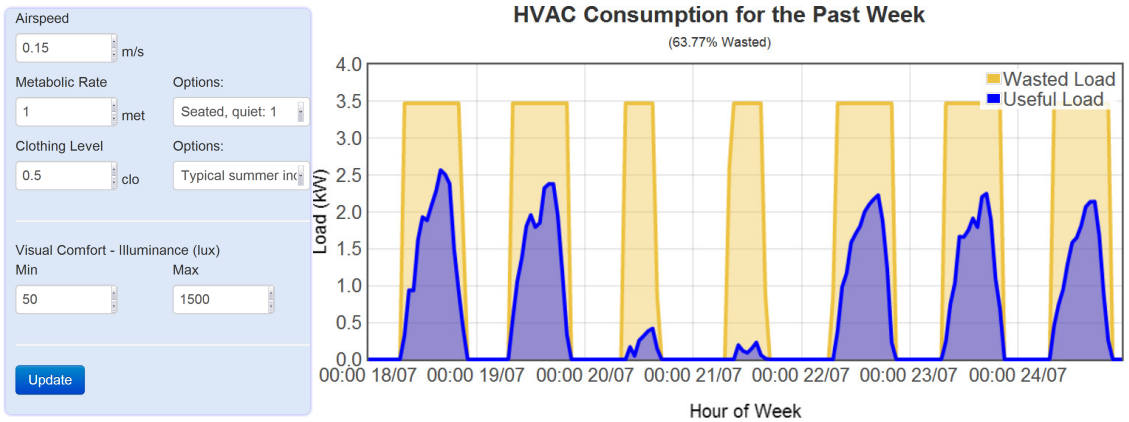


Figure 3.5: EnergyTrack user interface for HVAC consumption. Energy wasted is the difference between the actual and the useful consumption. The useful consumption is calculated using PPD parameters entered into the panel on the left.

consumed when the occupancy is zero. Such waste can be avoided if the last person to leave the office switches off the lights. For HVAC, wastage is given by the amount of energy that could be saved if the temperature setpoint were set so that at least 90% of the persons in the office were satisfied per the PPD metric. That savings can be achieved by simply increasing the temperature setpoint to the optimal value, which is 26.7°C for the ADSC testbed. We calculated that saving from the curve shown in Fig. 3.4. For computers, we define wastage by the amount of energy consumed when the computers are on but not performing any processing tasks. In such situations, they ought to be switched to standby mode.

3.4.3 Results of Comparative Evaluation

EnergyTrack and the baseline heuristic method were compared using the same set of data for a period of one month. The office staff were not informed about this energy wastage investigation, and thus their consumption behaviors should not have been influenced by the fact that they were being monitored during this period. The wastage estimates of EnergyTrack were significantly greater than those of the heuristic method, as shown in Table 3.1. The key reason for this difference is that EnergyTrack accounts for dynamic changes in occupancy and comfort, whereas the heuristic method does not. The table shows the wasted energy for a period of one month, in absolute terms and as a percentage of the

Table 3.1: Energy wastage estimates for one month (ADSC testbed)

Load	Heuristic Method	EnergyTrack
Lighting	458k Wh (15%)	2009 kWh (66%)
PC	442 kWh (14%)	1391 kWh (44%)
HVAC	400 kWh (32%)	770 kWh (62%)

total energy consumed by each appliance.

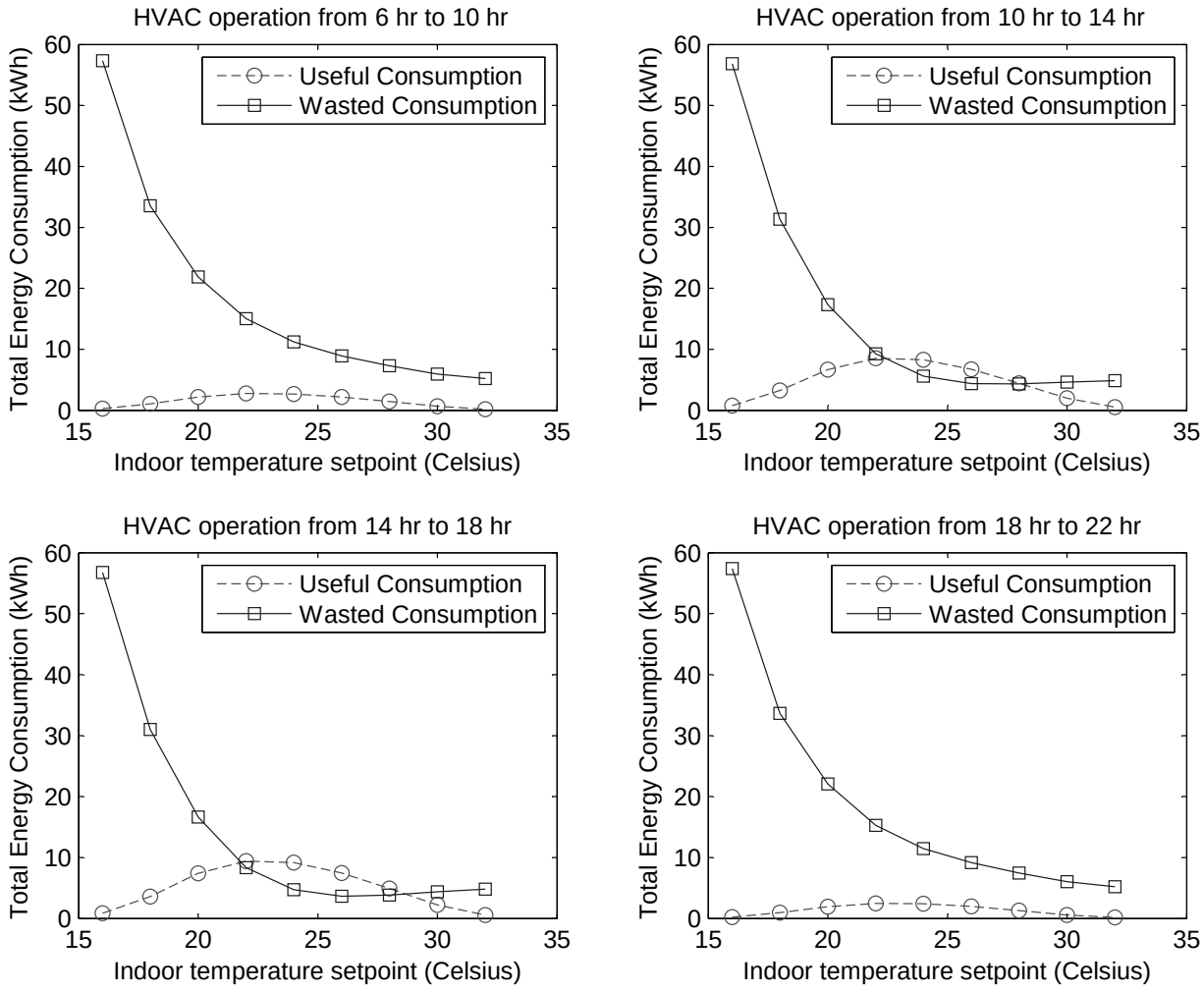


Figure 3.6: Energy use analysis of HVAC for different temperature setpoints over operation periods.

In addition to the comfort-cost trade-off depicted in Fig. 3.4, EnergyTrack incorporates occupancy and applies Eq. (3.1) to analyze the useful and wasted HVAC consumption for different temperature setpoints. These consumptions were analyzed for each 4-hour pe-

riod from 6:00 hrs to 22:00 hrs, as shown in Fig. 3.6. During the high-occupancy period (10:00–18:00 hrs), energy is most efficiently used at 24°C; at that temperature, the useful consumption is greater than the wasted consumption. However, during the low-occupancy period (6:00–10:00 hrs and 18:00–22:00 hrs), no satisfactory temperature setpoint exists, since the useful consumption is less than the wastage for all settings. That result is consistent with our intuition, since HVAC systems will be most efficient when the maximum number of occupants experience optimal thermal comfort.

For the one-month period, the heuristic method estimated the total energy wasted to be 1300 kWh, while EnergyTrack estimated it to be 4170 kWh. At an electricity tariff of 20 c/kWh, the advantage realized by the EnergyTrack method over the heuristic method is 574 dollars per month. That number effectively quantifies how much a facility manager would stand to gain from installing an automated control system that can realize the full savings potential estimated by EnergyTrack, as opposed to relying on manual control based on heuristic methods.

3.5 Evaluating the Demand Response Capacity of Buildings

The DR capacity of a building was introduced in Section 3.1. To determine that capacity, a large number of building-specific models and parameters need to be estimated. We divide those parameters into two categories. The first category, called *reference parameters*, includes the time of the week, occupancy at the time, and weather conditions. Those parameters cannot be manipulated by a building facility manager. The second category, *control parameters*, includes parameters that can be manipulated by a building facility manager in order to control consumption. The control parameters include the on/off state of various loads, the HVAC temperature setpoint, and the intensity of indoor lighting.

One could use simulation tools to precisely evaluate DR capacity under various settings of the aforementioned parameters. That, however, would require the detailed building specifications and in-depth domain knowledge needed to accurately model buildings. As a result, that approach does not scale well when applied to several buildings. We provide a scalable solution and a reliability measure that can be consistently applied across different buildings.

Our proposed measure captures the trade-off between providing a specified reduction in demand in response to DR signals, and the reliability with which that reduction can be achieved.

3.5.1 Demand Response Capacity Model

Reference Parameters

For the reference parameters, denoted by θ_r, we consider working or non-working days (θ_r^{wd}), hours of the day (θ_r^{hr}), occupancy level (θ_r^{occ}), external solar irradiance (θ_r^{sol}), and external temperature (θ_r^{temp}). Non-working days include Saturday, Sunday, and public holidays.

Control Parameters

For control parameters θ_c, we consider only two types of loads: HVAC and lighting. They are denoted by θ_c^{hvac} and θ_c^{light}, respectively. Unlike the reference parameters, the control parameters are defined separately for ADSC and ZEB in order to take account of their particular control capabilities.

For θ_c^{hvac}, we consider a temperature set-point control for ADSC and an ON/OFF control of the ventilation fans for the ZEB. The latter does not have setpoint control. For θ_c^{light}, we consider an ON/OFF control for ADSC and dimming capabilities for lights in the ZEB. The former does not have dimming control.

We assume that plug loads are not controllable for DR. The reason is that most plug loads in commercial buildings are electronic or electrical devices that need to be constantly running. While adjusting heating and lighting would be acceptable as long as minimum comfort levels are maintained, we assume that turning off plug loads would be unacceptable because that would result in disruption to the essential tasks of the commercial building occupants. The large variance in plug load demand causes a large variance in the total demand.

Table 3.2: Quantization of reference parameters

State	0	1	2	3	4	5
θ_r^{wd}	Sat, Sun, Public Hols	Mon–Friday	-	-	-	-
θ_r^{hr}	21–7	7–10	10–12	12–14	14–18	18–21
θ_r^{occ}	0	0–25	25–50	50–75	>75	–
θ_r^{sol}	0	0–200	200–400	400–600	>600	–
θ_r^{temp}	<21	21–24	24–27	27–30	>30	–

Note: the units for (θ_r^{hr}, θ_r^{occ}, θ_r^{sol},θ_r^{temp}) are (hrs, %, W/m^2, $^\circ C$)

Demand Model

We use D to denote the total demand of the building. D depends on the reference parameters θ_r and the control parameters θ_c. Let $\theta = (\theta_r, \theta_c)$ denote the vector of both the reference and control parameters, and let $D(\theta)$ denote the conditional random variable $D|\theta$. We wish to characterize the distribution of $D(\theta)$ with mean $\mu(\theta)$ and variance $\sigma^2(\theta)$ for different values of θ, but that is challenging because θ is a combination of both discrete and continuous parameters. Therefore, we quantize θ to make it discrete; and the quantization of the reference parameters θ_r is given in Table 3.2.

Figure 3.7 shows examples of the distribution of $D(\theta)$ for different values of θ_r, and constant default values of θ_c. Those examples show that $D(\theta)$ can be well approximated by a Gaussian distribution. That assumption was made in [58], but it did not help reveal any insights. Therefore, we do not make the assumption in this chapter.

3.5.2 Minimum Acceptable Demand

Let u refer to a vector of comfort metrics described in Section 3.3.2. In our study, the metrics include thermal and visual comfort, and they are dependent on the occupancy of the building because more occupants cause an increase in room temperature due to the release of body heat. Also, more occupants cause an increase in the obstruction of natural lighting sources (windows), requiring increased artificial lighting.

The minimum acceptable demand is the total demand corresponding to control settings θ_c, which are set such that the comfort metrics u are at the minimum acceptable level for

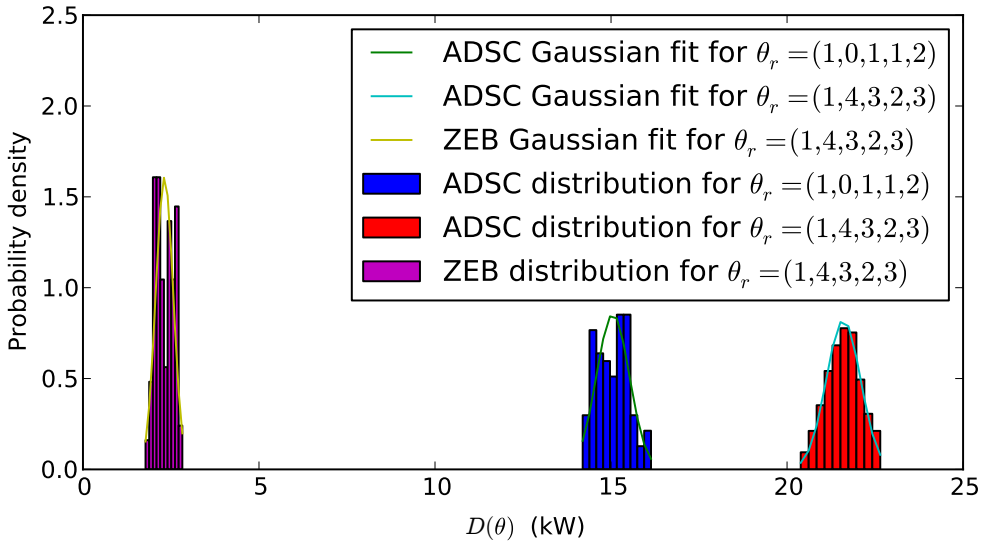


Figure 3.7: Conditional distributions of $D(\theta)$ for different values of $\theta_r = (\theta_r^{wd}, \theta_r^{hr}, \theta_r^{occ}, \theta_r^{sol}, \theta_r^{temp})$ and constant default values of θ_c in the ADSC and ZEB testbeds. The numbers in the parentheses are the column indices of Table 3.2 and map to values given in those columns.

the current occupancy $u_{min}(\theta_r^{occ})$.

$$\theta_c^* = \arg\min_{\theta_c} D(\theta_r, \theta_c) \tag{3.2}$$

$$\text{subject to } u \geq u_{min}(\theta_r^{occ}). \tag{3.3}$$

The minimum acceptable demand, $D(\theta_r, \theta_c^*)$, is the lowest demand at which the building facility manager will operate the building during a DR event that requires the building to reduce demand. The reduction of total demand as a result of setting the controllable parameters to θ_c^* is given as follows.

$$\Delta D(\theta_r, \theta_c) = D(\theta_r, \theta_c) - D(\theta_r, \theta_c^*). \tag{3.4}$$

The reduction $\Delta D(\theta_r) \geq 0$ as $D(\theta_r, \theta_c^*)$ is the minimum acceptable demand. The total

Table 3.3: Minimum acceptable demand required given occupancy states

θ_r^{occ}		0	1	2	3	4
Temperature Setpoint	$^\circ C$	30	27	27	26	26
ADSC	θ_c^{hvac*}	30	27	27	26	26
	$D_{hvac}(\theta_r, \theta_c^{hvac*})(kW)$	0	0	0	2.778	2.778
ZEB	θ_c^{hvac*}	OFF	OFF	OFF	ON	ON
	$D_{hvac}(\theta_r, \theta_c^{hvac*})(kW)$	0	0	0	1.5	1.5
Lighting Setpoint	Lux	0	500	500	600	600
ADSC	θ_c^{light*}	OFF	OFF	OFF	OFF	OFF
	$D_{light}(\theta_r, \theta_c^{light*})(kW)$	0	0	0	0	0
ZEB	θ_c^{light*}	OFF	OFF	OFF	600	600
	$D_{light}(\theta_r, \theta_c^{light*})(kW)$	0	0	0	0.136	0.136

Note: values are based on (θ_r^{wd},θ_r^{hr}, θ_r^{sol},θ_r^{temp}) = (1,4,2,3) from Table 3.2.

demand can be split into its components as follows.

$$D(\theta_r, \theta_c) = D_{plug}(\theta_r) + D_{hvac}(\theta_r, \theta_c^{hvac}) + D_{light}(\theta_r, \theta_c^{light}) \tag{3.5}$$

$$D(\theta_r, \theta_c^*) = D_{plug}(\theta_r) + D_{hvac}(\theta_r, \theta_c^{hvac*}) + D_{light}(\theta_r, \theta_c^{light*}), \tag{3.6}$$

where the plug load D_{plug} is not dependent on the control parameters, but the HVAC load D_{hvac} and light load D_{light} are dependent on their respective control parameters.

Table 3.3 contains the values of θ_c^{hvac*}, θ_c^{light*}, and the corresponding values of D_{hvac} and D_{light}. The occupancy levels θ_r^{occ} in the table are given by the column indices to Table 3.2. Note that θ_c^{hvac*} has setpoint control for ADSC, but only ON/OFF control for ZEB. Similarly, θ_c^{light*} has dimmable control for ZEB, but only ON/OFF control for ADSC.

3.5.3 Demand Response Capacity Analysis

DR is usually associated with compensation to consumers for enduring the inconvenience caused by the reduction of demand $\Delta D(\theta_r)$. That compensation is usually on a dollar-per-kilowatt measure. Therefore, it is essential for utilities to be able to ensure that the reduction in demand is being honored.

An unscrupulous consumer could start with a higher consumption level $\Delta D(\theta_r, \theta_c)$ just in order to monetize on a larger $\Delta D(\theta_r)$ during a DR event. To avoid such a circumstance, the

Table 3.4: Look-up table example for parameters and total consumption statistics. During a DR event, the reduction is made from these values as per an agreement with the utility.

$(\theta_r^{wd}, \theta_r^{hr}, \theta_r^{occ}, \theta_r^{sol}, \theta_r^{temp})$	θ_c^{light}	θ_c^{hvac}	$\mu(\theta_r, \theta_c)$	$\sigma(\theta_r, \theta_c)$
$(1, 4, 2, 2, 3)$	ON	$23°C$	1.2 kW	0.2 kW
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

consumer and the utility can agree on reasonable control parameters θ_c that are appropriate for a given θ_r in the absence of a DR event. We propose the construction of a look-up table for each testbed site that stores the agreed-upon control parameters $(\theta_c^{light}, \theta_c^{hvac})$ for the reference parameters in the absence of a DR event. It also stores the mean $\mu(\theta_r, \theta_c)$ and standard deviation $\sigma(\theta_r, \theta_c)$ of the total load $D(\theta_r, \theta_c)$ corresponding to those parameters. An example of such a look-up table is given in Table 3.4.

During a DR event, the utility uses $\mu(\theta_r, \theta_c)$ as a baseline against which the reduction is measured, because the utility and consumer previously agreed upon that value as the expected demand $E[D(\theta_r, \theta_c)]$. Thus, the reduction that the utility would use is given by $\Delta D(\theta_r, \theta_c)$ and expressed as follows.

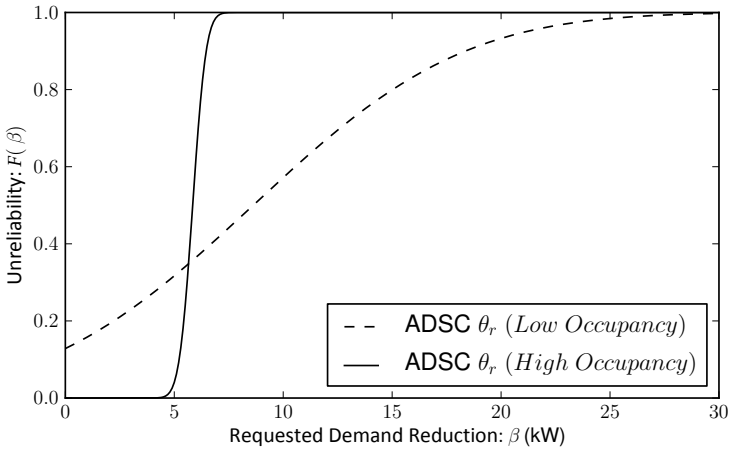
$$\Delta\delta(\theta_r, \theta_c) = \mu(\theta_r, \theta_c) - D(\theta_r, \theta_c^*). \tag{3.7}$$

Note that $\Delta\delta(\theta_r, \theta_c)$ is different from $\Delta D(\theta_r, \theta_c)$, which was defined in Eq. (3.4). $\Delta\delta(\theta_r, \theta_c) \neq \Delta D(\theta_r, \theta_c)$ at specific time instants, but on average, they are equal. Use of $\Delta\delta(\theta_r, \theta_c)$ obviates the need for the utility to constantly verify the integrity of $D(\theta_r, \theta_c)$, because its expectation $\mu(\theta_r, \theta_c)$ has been agreed upon. Thus, there is no room for fraudulently claiming larger DR credits.

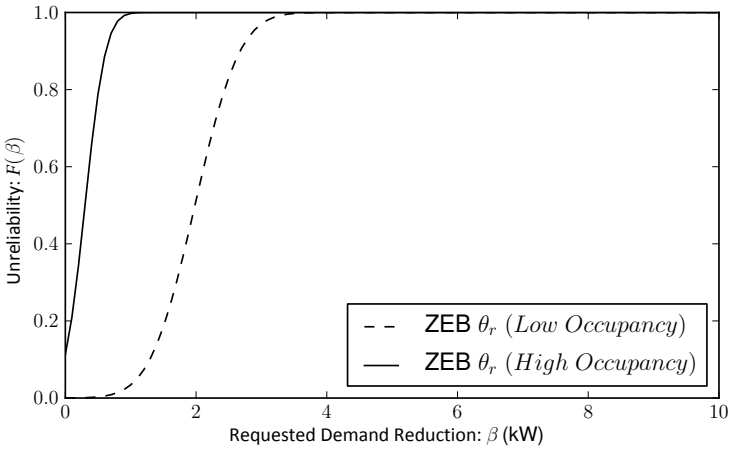
Let β denote the DR capacity of a building, defined as the reduction of demand that a consumer can provide with a certain reliability. We define the uncertainty of DR as follows.

$$F_\theta(\beta) = P(\Delta\delta(\theta_r, \theta_c) \leq \beta), \tag{3.8}$$

where $F_\theta(\beta)$ is the cumulative distribution function (CDF) of $\Delta\delta(\theta_r, \theta_c)$. $F_\theta(\beta)$ is monotonically increasing, by properties of the CDF. This formula for the uncertainty of DR is



(a) ADSC testbed



(b) ZEB testbed

Figure 3.8: Trade-off between uncertainty and DR capacity for a DR duration of 1 hour, where $\theta_r = (1, 4, \theta_r^{occ}, 2, 3)$ and $\theta_r^{occ} = 1$ (low) or 3 (high).

consistent with the definition of reliability in terms of the CDF of a failure rate in probabilistic modeling theory.

3.5.4 Evaluation Results of Demand Response Capacity

In Fig. 3.8 we compare the DR capacities of the ADSC and ZEB testbeds for different reference parameter settings θ_r, with a demand response period of 1 hour. For each testbed we compute the trade-off between uncertainty and DR capacity for $\theta_r^{occ} = 1, 3$ (i.e. low and high occupancy states), while the rest of the reference parameters are fixed at θ_r^{wd}=1 (i.e.,

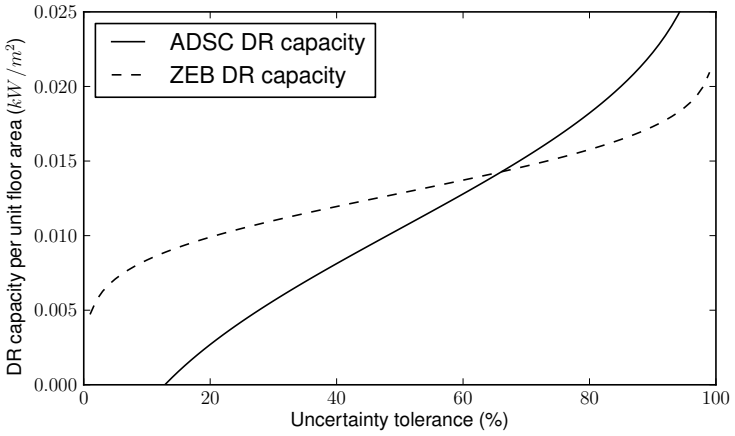


Figure 3.9: DR capacity comparison between ADSC and ZEB at low occupancy given $\theta_r = (1, 4, 1, 2, 3)$ for a duration of 1 hour.

working day), θ_r^{hr}=4 (i.e. 2:00–6:00 P.M.), θ_r^{sol}=1 (i.e., sunny to partly cloudy), and θ_r^{temp}=3 (i.e., warm weather).

For the same uncertainty level, we expect greater demand reduction capacity when there is low occupancy because the demand is usually higher than it needs to be when there are fewer people. In other words, there is more energy wastage in low-occupancy periods, so there is greater potential for reducing energy use. At high occupancy, the loads are better utilized, so there is less potential for demand reduction. For the ZEB, that trend holds for all uncertainty levels, as seen in Fig. 3.8(b). For ADSC, however, that trend holds only until the demand reduction capacity rises to about 6 kW, as seen in Fig. 3.8(a). For smaller demand reduction capacities, the uncertainty is higher during low-occupancy periods in ADSC. The reason is that the variance of the load $\sigma(\theta)$ is much higher during low occupancy than it is during high occupancy. That large variance during low-occupancy periods, which can be seen by comparing Fig. 3.1(b) and Fig. 3.2(b), explains the lack of certainty with which demand can be reduced. It makes it more likely that ADSC can provide high DR capacity (over 25 kW) during low-occupancy periods.

In Fig. 3.9 we compare the DR capacities of ADSC and the ZEB after the values have been normalized with respect to their floor area. It is clear from the figure that the ZEB has greater DR capacity at low and moderate uncertainty tolerance values ($< 60\%$). That is likely due to the presence of more precise control capabilities, i.e., dimmable lighting

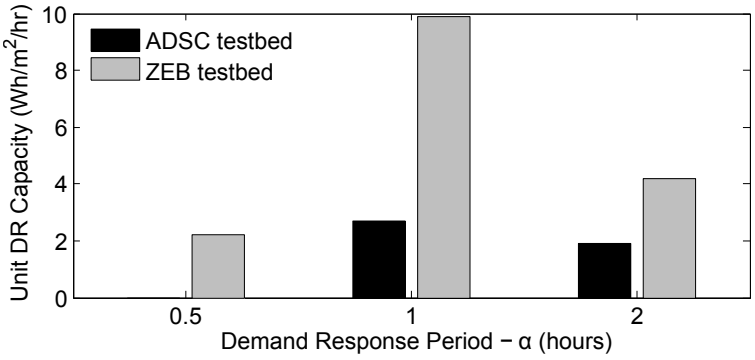


Figure 3.10: DR capacity comparison for different demand response periods, given $\theta_r = (1, 4, 1, 2, 3)$.

(see Fig. 3.3). When the uncertainty tolerance exceeds 60%, ADSC does have a larger DR capacity than the ZEB; however, such high uncertainty about DR performance would likely be undesirable for building owners and/or utilities. That is particularly true in market settings that enforce noncompliance penalties when DR that promised is not delivered [66].

Finally, in Fig. 3.10 we compare demand response capacity for different DR periods of 0.5 hour, 1 hour, and 2 hours for the ADSC and ZEB testbeds. For a fair comparison, we normalized the DR capacities with respect to the testbeds' floor areas for each choice of DR period. We refer to this as *unit DR capacity*. The figure shows that both the ZEB and ADSC have the highest unit DR capacity when the DR duration is 1 hour. In particular, for ADSC no DR capacity is available if the DR period is less than half an hour because the expectation of the demand $E[D(\theta_r, \theta_c)]$ may not be equal to the pre-agreed expectation $\mu(\theta_r, \theta_c)$ over such short periods of time because of high variance. For both HVAC and lighting, a longer demand response period makes it difficult to ensure that the reference parameters (such as occupancy and weather) will remain constant over that period. Thus, the capacity decreases with increase in duration, and consequent variability in $D(\theta_r, \theta_c)$. These are useful results that highlight the importance of choosing the right DR period for a building's unique consumption and occupancy patterns.

3.6 Related Work

There are two stages of demand response capacity evaluation for buildings: the audit stage [67], and the measurement and verification (M&V) stage [68]. An audit takes place before the site enrolls in a DR program; its purpose is to identify the amount of responsive load at the site, and the control actions that will be taken in an event. The M&V stage occurs after a DR event, and allows aggregators or market authorities to confirm that a specified curtailment did indeed occur. In this work, we are concerned with the demand response audit stage.

Demand response audits [67] generally involve a high degree of human effort. Within the last decade, researchers have adopted top-down approaches for characterizing typical DR actions and audit procedures. For example, the authors of [69] analyze utility meter data from dozens of buildings and propose methods to help facility managers identify demand response opportunities. Following a similar direction, [70] provides a high-level overview of demand response strategies for commercial buildings. Simulation tools have also been developed to allow building owners to estimate their demand response potential based on a variety of site-specific inputs and typical control strategies [71].

At a more detailed level, a growing body of work has emerged that examines specific building loads for demand response purposes. Such efforts often include the use of sensor networks in the built environment. In this area, much attention has been devoted to plug loads or miscellaneous power [72, 73], including specific sources of plug load such as laptops [74]. Outside of plug loads, the other dominant building energy end-uses tend to be lighting and HVAC. For lighting and HVAC systems, researchers have studied optimal control methods that use measurements from sensor networks [75–77].

Our work differs from the above efforts in several important ways. Since we use indoor sensor networks rather than utility metering data, we are able to provide a more detailed view of building energy consumption and demand response potential than can top-down audit approaches [69, 71]. Although several related efforts (e.g., [73, 75]) have also leveraged wireless sensor networks, our scope is more holistic in that it includes multiple electricity end-uses (HVAC, lighting, and plug load) as well as consideration for occupants' thermal

and visual comfort.

3.7 Conclusion

In this chapter, we presented EnergyTrack, a system that analyzes and interprets energy consumption patterns in buildings. We propose an analysis model for energy usage that jointly considers occupancy levels and the utility provided by end-loads. We demonstrated the application of EnergyTrack for energy use analysis in two real testbeds.

We presented a framework for evaluating the demand response capacity of buildings. In doing so, we adopted a data-driven approach that makes no assumptions about physical or regression models of the building's power consumption. We used a look-up table that associates observed or projected consumption with a particular set of conditions (or rules) learned from our dataset. We deployed sensor networks in two testbeds, collected data from the sensors over a period of 10 weeks, and used our framework to evaluate the DR capacity for these testbeds.

In conclusion, we demonstrated the claim in our thesis statement in the context of utilizing demand response as an energy resource. We showed that an empirical model constructed using statistical methods outperformed a heuristic model by three times in its ability to maximize demand reduction while respecting occupant comfort requirements. We also used statistical methods to quantify the uncertainty associated with demand reduction. That empowers utilities to choose buildings that can provide the required demand reduction with high certainty. The work in this chapter was peer-reviewed and published in [58] and [59].

CHAPTER 4

FRAMEWORK FOR IDENTIFYING AND LOCALIZING METER FRAUD

"There's no such thing as a foolproof system. That idea fails to take into account the creativity of fools."

– Frank Abagnale Jr.

Having discussed ways in which energy resource utilization can be improved in the context of wind generation (Chapter 2) and building energy use (Chapter 3), we now discuss how that improvement can be undermined if the data have been compromised. In particular, we consider attacks that compromise the Advanced Metering Infrastructure (AMI) for monetary benefit through meter fraud. Instead of providing demand reduction as a resource (as described in Chapter 3), the attacker would falsely claim, by compromising meter readings, that they have reduced demand. In reality, the attacker would not have reduced demand, and would have instead artificially reduced his or her own electricity bill. Similarly generators can compromise their meter readings and claim to generate more than they are actually generating so that they get paid more. That behavior clearly negates the benefits of improving utilization through the legitimate increase in generation or decrease in demand. We address that issue through the exploration of fraud detection in this chapter and in Chapters 5–6.

To provide some context to the problem described in this chapter, we describe AMI-related security issues more broadly. The AMI provides a means for electric utilities to monitor the electric distribution grid. Vulnerabilities in that infrastructure could allow cyber adversaries to compromise the grid in ways that can have adverse effects on utilities and consumers. These effects include electricity theft, disruption of electricity service, and damage to the electricity delivery infrastructure. Although one should defend smart grids against a diversity of possible attacks (as described in [78]), we focus our attention on electricity theft, which is one of the most important problems faced by electricity suppliers and utilities around the world.

Bloomberg News reported that electricity theft in India contributes to blackouts and costs

$17 billion in lost revenue annually [79]. According to the World Bank, electricity theft contributes to a loss in electricity delivery of over 25% of generated supply in India, 16% in Brazil, 6% in China and the U.S., and 5% in Australia [79]. Theft in these countries is almost always achieved by tapping into electric distribution lines. To detect these thefts, utility companies such as BC Hydro have been convincing consumers to install smart meters [8,80]. However, there has been some push-back as consumers have begun to realize that smart meters are vulnerable to cyber intrusions [81]. In 2010, the Cyber Intelligence Section of the FBI reported that smart meter consumptions were being under-reported in Puerto Rico, leading to annual losses for the utility estimated at $400 million [10]. In 2014, BBC News reported that smart meters in Spain were hacked to cut power bills [9]. Given that smart meters can be compromised, the smart meter roll-out efforts of utilities such as BC Hydro may only increase the attack surface for cyber-intrusion-based theft methods.

4.1 Summary of Contributions

In this chapter, we present *F-DETA*, a framework for systematically identifying, classifying, and detecting electricity theft attacks targeted at utilities, consumers, and DERs. Specifically, we classify an attack based on (1) its ability to evade detection methods currently used in industry and (2) its applicability in various electricity pricing schemes. Furthermore, the fundamental nature of our approach has allowed us to identify seven classes of attacks, only two of which have been presented in related work. Some of these classes may distribute the monetary loss across consumers, at no loss to the utility. All the attack classes identified in this chapter can be launched by compromising the integrity of smart grid communication signals (e.g., price or consumption measurements). Only some of these attacks may also be achieved by tapping electric distribution lines. The identification and analysis of these attack classes guides the creation of our detection approaches that dramatically mitigate, if not completely eliminate, electricity theft.

The rest of this chapter is organized as follows. The preliminaries in Section 4.2 provide background to understanding the attack model presented in Section 4.3. A topological representation of the electric distribution system is described in Section 4.4. That tree-

based representation guides the formal description of attack strategies in electricity theft. The classification of those attack strategies is presented in Section 4.5. The framework, adapted to the context of DER fraud, is presented in Section 4.6. The detection model, along with its associated assumptions, is presented in Section 4.7. We discuss related work in Section 4.8.

4.2 Preliminaries

We analyze changes in electricity consumption, price, and their reported values in discrete time. In our power network model, all electricity consumers have meters installed to measure their consumption. These meters are all electronic with network interfaces, not traditional analog meters. We assume that these *smart* meters have a fixed polling time period (Δt) that we treat as our basic discrete time unit. We label each time period using an integer $t \in \mathbb{Z}_{\geq 0}$. The smart meter readings in our model represent the average demand during each time period t. The average demand can be multiplied by Δt to obtain the consumption for billing purposes.

We use $D_C(t)$ to denote the *average demand* of a consumer C during the time period t, where $D_C(t) \in \mathbb{R}$. For clarity, we implicitly assume that $D_C(t) \geq 0$, unless otherwise stated. When $D_C(t) < 0$, we say the consumer is generating electricity at time t, and our framework is equally valid in this distributed energy resources (DER) context. $D'_C(t)$ denotes the average demand *reported* by the smart meter to the utility during time t. Under that notation, $D_C(t) \neq D'_C(t)$ would imply that the smart meters (or network communications) have been compromised or are malfunctioning. Unless otherwise stated, we assume that smart meters are correctly functioning, so the inequality would imply that they have been compromised.

To the best of our knowledge, we are the first to take into account electricity pricing schemes when studying electricity theft attack strategies. It is important to consider pricing schemes, because false data injections in a certain service area (or against a certain set of consumers) could be tailored to the specific pricing scheme in that area (or pricing scheme adopted by those consumers). We consider flat-rate, time-of-use, and real-time pric-

ing schemes. In *flat-rate pricing*, the price stays constant throughout a billing cycle spanning T time periods. This is the traditional pricing scheme and is prevalent all over the world. In *time-of-use* (TOU) pricing, the price of electricity varies according to a plan. Certain hours of the day are designated as *peak*, *partial-peak*, or *off-peak*, and the prices for these hours are published by the utility before consumers agree to sign up for this scheme. Finally, in *real-time pricing* (RTP), the prices change in a nondeterministic manner that captures the dynamic market trends in electricity demand and supply.

Let $\lambda(t)$ denote the electricity price during the time period t, where $\lambda(t) \in \mathbb{R}_{\geq 0}$. Note that the price does not necessarily change between smart meter polling periods, and that price updates are usually less frequent than polling reports. We assume that the price update period is $k\Delta t$, where $k \in \mathbb{Z}_{>0}$. For simplicity, we assume that the price of electricity consumed is the same as the price of electricity generated by a consumer in a distributed generation setting, but the framework can be extended to account for differences in those prices.

4.3 Attack Model

The attacker in this chapter is an electricity consumer in the electric distribution grid who consumes more electric energy than she pays for (or generates less than she is paid for). We follow network security naming convention and refer to her as Mallory, the attacker whose intentions are malicious. Mallory's intention in stealing electricity is to make a monetary profit at the expense of the utility or her neighbors. Broadly speaking, an attacker can steal electricity by manually tapping electric power lines or by electronically injecting false readings. Our focus is on the latter method, which assumes that *either the smart meter or the communication link has been compromised, and the attacker is now an insider in the system*. This is reasonable to assume given the evidence of real hacking incidents in [10, 81] and [9].

In the attack model, we assume that the smart meters are in either of two states: compromised or correctly functioning. If meters are faulty without the intervention of an attacker, it is possible that electricity can be unaccounted for. However, this can be easily detected and

investigated, as we later show in Subsection 4.4.2. Therefore, our attack model is concerned only with faults that are intentionally introduced by an attacker.

If the billing cycle contains T time periods, then a single attacker A can successfully steal electricity (in other words, execute an electricity theft attack) if and only if the following condition holds:

$$\sum_{t=1}^{T} \lambda(t)[D_A(t) - D'_A(t)] > 0. \tag{4.1}$$

This is a simplified form of the expression that describes Mallory's profit or monetary advantage α, which is given by the difference between what the utility should bill her based on actual consumption, B_{Utility}, and what the utility actually bills her based on fraudulently reported consumption, B'_{Utility}:

$$\begin{aligned} \alpha &\triangleq B_{\text{Utility}} - B'_{\text{Utility}} \\ &= \sum_{t=1}^{T} \lambda(t) D_A(t) \Delta t - \sum_{t=1}^{T} \lambda(t) D'_A(t) \Delta t \\ &> 0. \end{aligned} \tag{4.2}$$

Here the units may be given as follows: λ is in \$/kWh, D is in kW, Δt is in hours, and α is in \$ (dollars). Since Δt is a positive constant, it does not factor into Eq. (4.1). Mallory's objective is to maximize α subject to the constraint that her attack must go undetected.

Our principled approach to searching for attack strategies begins with the observation that it is necessary for Mallory to under-report her consumption at some time t. This is formally stated in the following proposition:

Theorem 1. *Under any electricity pricing scheme with positive prices, in order to make a profit, an attacker must under-report her readings in at least one time period. In mathematical terms, if constraint Eq. (4.1) holds, then $\exists t$ such that $D'_A(t) < D_A(t)$.*

Proof (by contradiction): Assume $\forall t, D'_A(t) \geq D_A(t)$, then:

$$\sum_{t=1}^{T} \lambda(t)[D_A(t) - D'_A(t)] \leq 0, \tag{4.3}$$

which contradicts Eq. (4.1). ■

Therefore, any strategy that does not under-report the demand at any time cannot be an electricity theft attack. Under-reporting can be achieved either by compromising the integrity of the smart meter readings or by tapping the electric power line immediately upstream of the smart meter. In the latter case, the smart meter measures only what is downstream of it and misses what was tapped out upstream of it. As a result, the meter, though not compromised, reports values that are lower than what is actually consumed. Theorem 1 also holds when Mallory is generating electricity, as under-reporting consumption is equivalent to over-reporting generation.

4.4 Electric Distribution Grid Topology Representation and the Balance Check

Certain attacks can be launched and detected using information about the electric distribution grid topology. Most of these topologies in practice are radial, so we only assume radial topologies in this chapter. A radial topology can be represented as an unbalanced n-ary tree, where n represents the maximum number of consumers, or leaf nodes, connected to a single node. Another common topology is the loop system, which was designed to improve the reliability of power delivery. Loop systems are essentially radial, as the loop is closed only during a fault (see [82]). As a result, power to a consumer at any one time is supplied through a single path from the distribution substation, which we refer to as the *root node* of the n-ary tree. Through a series of transformers and protective equipment, power is supplied from the root node to the leaf nodes. The root node would typically lie in a substation that connects the transmission (high-voltage) electric grid with the distribution (low-voltage) electric grid.

Since active power is additive, the total average power that is supplied at a node in the tree at time t is equal to the sum of the average demands at all its child nodes at time t. This is illustrated in Fig. 4.1, where network losses are also modeled as leaf nodes. Figure 4.1 helps explain the balance check, which is an important detection mechanism used in industry.

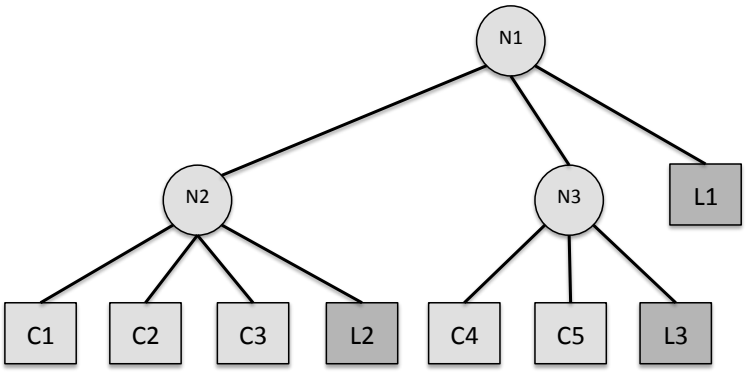


Figure 4.1: Illustration of a radial power network topology as an n-ary tree. Circles represent internal nodes $N1$–$N3$. Squares represent leaf nodes that include end-consumers $C1$–$C5$ and network losses $L1$–$L3$. In this example, $D_{N1}(t) = D_{N2}(t) + D_{N3}(t) + D_{L1}(t)$ and $D_{N3}(t) = D_{C4}(t) + D_{C5}(t) + D_{L3}(t)$.

4.4.1 The Balance Check

Let $D_N(t)$ represent the demand at internal node N at time t. In addition, let C/L be the set of all consumer/loss nodes that are descendants of N. The loss nodes model line impedances and transformer losses. Then, the following equation holds:

$$D_N(t) = \sum_{c \in C} D_c(t) + \sum_{l \in L} D_l(t). \tag{4.4}$$

This equation is the first step to arriving at the balance check, and is the foundation for a recent patent [83]. The balance check is used in industry to check whether readings add up correctly. The check is performed at internal nodes in the distribution grid topology, which we call *balance meters*.

In [83], D'_N, which is the balance meter reading at N, is compared with the readings reported by smart meters located at each $c \in C$. Utilities and the authors of [83] assume that there is no electricity theft if the following condition is satisfied:

$$D'_N(t) = \sum_{c \in C} D'_c(t) + \sum_{l \in L} D_l(t). \tag{4.5}$$

In this chapter, we show that the above assumption is false, and identify attacks that are effective despite having satisfied the condition. Note that the losses are not reported, but

calculated by utilities based on known values of distribution system component specifications, such as line impedances. That calculation is discussed in [84]. Thus, we do not require a symbol like D'_l for reported loss values.

Assuming that these balance meters are trusted, the reported meter measurement at node N, which is D'_N, is equal to D_N. Thus, we can combine Eq. (4.4) and Eq. (4.5) to obtain the simplified balance check for electricity theft detection:

$$\sum_{c \in C} D'_c(t) = \sum_{c \in C} D_c(t). \tag{4.6}$$

4.4.2 Detecting Faulty or Compromised Meters

Let W denote the event that *a meter at a specified node reports a balance check failure*. The following points would help identify a faulty or compromised meter.

If W is true for an internal node, it must be true for all its ancestors, all the way up to the root node. An alarm should be raised for investigation if W is true for an internal node and false for its immediate parent node. Such a situation would imply that at least one of the two meters is faulty or that at least one of them has been compromised. Note that the inverse of this implication is not true. If W is false for a node, then it is still possible that W is true for the node's parent. This would imply that W is true for at least one of the parent's other child nodes.

If a parent of internal nodes has W as true, and all of its child nodes have W as false, then an alarm should be raised for investigation. This situation implies that at least one of the children, or the parent itself, is faulty or compromised.

4.4.3 Investigating Failure of the Balance Check

If the balance check fails at a balance meter that is correctly functioning and has not been compromised, then it implies that electricity has been stolen. If the balance meter were malfunctioning or compromised, the failure of a balance check might not indicate electricity theft. Even so, the meter should be investigated and fixed by the utility. Such investigations,

Table 4.1: Attack classification

Attack Class	1A	2A	3A	1B	2B	3B	4B
Possible despite Balance Check	N	N	N	Y	Y	Y	Y
Possible with Flat Rate Pricing	Y	Y	N	Y	Y	N	N
Possible with TOU Pricing	Y	Y	Y	Y	Y	Y	N
Possible with RTP	Y	Y	Y	Y	Y	Y	Y
Requires ADR	N	N	N	N	N	N	Y

currently made periodically [85], can be made less frequently, and in a systematic manner. In finding the faulty meter, the worst-case search order is $O(N)$. If the following approaches that exploit the tree structure of the topology are used, the effort and investment incurred by the utility can be minimized.

Case 1 (every internal node has been instrumented with a meter): Finding the deepest meter in the tree that reports a failure of the balance check would identify the geographic neighborhood that needs to be investigated. If all the internal nodes are trusted and functioning correctly, then the deepest internal nodes to report the failure would have a limited number of consumer leaf nodes connected to them. These leaf nodes would then need to be manually inspected. One or more of these consumers must be an attacker.

Case 2 (at least one internal node has not been instrumented with a meter): In this case, the utility could send a serviceperson with a portable meter and then perform a search similar to the breadth-first search tree traversal algorithm. Starting with the root node, the serviceperson would check each child of the node to see if the readings match the readings of the smart meters of consumer nodes that are descendants of the child, accounting for the nodes that represent losses due to impedance in the tree representation. After each check, he or she would investigate only the subtree of the node whose check failed. The other subtrees would not need to be investigated.

While it is true that the failure of the balance check implies that investigation is needed, it would be false to say that if the balance check is satisfied, then there has been no theft. Unfortunately, that was not explicitly recognized in [83] and other related work. In the next section, we show that there are theft attacks that can circumvent the balance checks.

4.5 Classification of Attacks

Our primary classification of attack strategies is based on those that fail the balance check (*Attack Classes 1A–3A*), and those that successfully circumvent it (*Attack Classes 1B–4B*). We identify seven classes of attacks in this chapter, and summarize them in Table 4.1, based on whether they are possible under different pricing schemes, and whether they require automated demand response (ADR) to be in place. We now define the attack classes, and will later describe ADR in the context of Attack Class 4B. We hypothesize that electricity theft attacks in practice may be a combination of one or more of these seven attack classes.

4.5.1 Attacks that Can be Detected using Balance Checks

In this subsection, we describe theft Attack Classes 1A, 2A, and 3A, for which the balance check in Eq. (4.6) can be used to locate where the theft is occurring within the power network. As acknowledged in [86], this method does not identify the individual attacker, but it dramatically reduces the search space for the investigation, which can then be done manually. Although the simplified balance check in Eq. (4.6) works against these attacks, it is possible for the balance meters to be compromised. Mallory, at a leaf node of the tree, would only need to compromise the balance check meters in the direct route to the root node. This direct route is the depth of the branch of the tree that supplies Mallory. The tree depths, which we have seen in the distribution grid models that we have used in previous work, range from 5 to 135. For an unbalanced tree, the number of meters that Mallory needs to compromise would be $O(N)$ in the worst case, where the tree is a linear structure. If the tree were balanced, the number of meters she needs to compromise would be $O(\log(N))$.

Attacks under Flat-rate Pricing Schemes

In our attack model, under the flat-rate pricing system, smart meter consumption readings reported to the utility are compromised, while pricing signals are not compromised. The assumption that pricing signals are not compromised is reasonable under a flat-rate pricing scheme, since price signals are fixed and pre-decided. Thus, any deviation from the fixed

price is suspicious and should be investigated.

The difference between what Mallory pays and what she should pay, as given in Eq. (4.1), quantifies the loss for the utility during the billing cycle. Depending on how the utility business model is structured, the price of the stolen electricity is either paid by the utility itself or jointly paid as service fees by all the consumers in the system.

Under the flat-rate pricing scheme assumed in [86], the attack condition Eq. (4.1) is reduced to the following:

$$\sum_{t=1}^{T} D'(t) < \sum_{t=1}^{T} D(t). \tag{4.7}$$

Within this scheme, the attack can take either of two approaches:

Attack Class 1A: The LHS of Eq. (4.7) is not varied, in which case the reported measurements $D'(t)$ do not deviate from Mallory's typical consumption. Instead, Mallory consumes more electricity than is typical, so the RHS of Eq. (4.7) is increased. This attack is potentially limited only by the capacity of the power network to meet Mallory's demand. Therefore, a large amount of electricity can potentially be stolen.

Attack Class 2A: The RHS of Eq. (4.7) is not varied, in which case Mallory does not change her typical behavior. Instead, the LHS of Eq. (4.7) is artificially decreased in order to satisfy the condition. This strategy is the only attack strategy that is presented in [86], and its severity is more limited than that of Attack Class 1A.

To quantify the limited amount of electricity that can be stolen under Attack Class 2A, we define a threshold $\tau \geq 0$ below which reported consumptions $D'(t)$ can be correctly classified as a theft attack. As an example in [86], τ can be defined as the minimum of daily consumption averages over a fixed number of days. The smallest value that τ can possibly take is 0. Therefore, the upper bound of the electricity that Mallory can steal is her typical consumption.

All the detection algorithms proposed in the fraud detection section of this dissertation, and in [86], are based on analyzing the change in the pattern of reported smart meter readings under the attack. Under Attack Class 1A, there is no change in the reported readings pattern, and therefore the attack would go completely undetected. However, it can be detected with the balance check.

Attacks under Time-of-Use and Real-Time Pricing Schemes

Under variable pricing, we default to Eq. (4.1), since $\lambda(t)$ is not constant. Attack Classes 1A & 2A are still possible in this context. In addition to those attacks, a third attack is possible, and it is described next.

Attack Class 3A: In this attack, Mallory does not steal any electricity, but shifts her load in such a way that she still makes a monetary profit. As indicated in Table 4.1, this attack is not possible with a flat-rate structure, as it requires Mallory to report that her consumption happened at a time when the price was low, when it actually happened when the price was high.

We now formalize Attack Class 3A. Consider two time periods t_1 and t_2 such that $\lambda(t_1) < \lambda(t_2)$. This can happen when t_1 is an off-peak period and t_2 is a peak period in a time-of-use (TOU) pricing system. At time t_1, Mallory may over-report her off-peak electricity consumption such that the difference $D'_A(t_1) - D_A(t_1) > 0$. Similarly, she may under-report her consumption during the peak period so that the difference $D_A(t_2) - D'_A(t_2) > 0$. If the differences match, then $D'_A(t_1) + D'_A(t_2) = D_A(t_1) + D_A(t_2)$, so the total demand is equal to the reported demand and no electricity is stolen. Even so, by making it appear as though Mallory has shifted her load from the peak to the off-peak period, she can profit without having stolen any electricity. Although we have used TOU pricing to illustrate this attack, equivalent arguments can be made for real-time pricing.

4.5.2 Attacks that Circumvent Balance Checks

In this subsection, we identify classes of attacks that go undetected by balance meter checks. Consider an electric distribution network node to which $M + 1$ consumers are connected: Mallory A and a set of M innocent neighbors $N = \{N_1, N_2, ..., N_M\}$. Physically, that node may be a bus or a transformer. The balance check constraint at that node at time period t is an instance of Eq. (4.6) given by:

$$D_A(t) + \sum_{n \in N} D_n(t) = D'_A(t) + \sum_{n \in N} D'_n(t). \tag{4.8}$$

The following theorem helps us identify theft strategies that meet the above constraint:

Theorem 2. *In order to make a profit while circumventing the balance check, an attacker must over-report the readings of at least one of her neighbors. In mathematical terms, if both constraints Eq. (4.1) and Eq. (4.8) hold, then* $\exists(n,t)$ *such that* $D'_n(t) > D_n(t)$ *where* $n \in N$.

Proof (by contradiction): Given that Eq. (4.1) holds, we know from Theorem 1 that $\exists t$ such that $D'_A(t) < D_A(t)$. For every such t, if we can prove that $\exists n$ such that $D'_n(t) > D_n(t)$, then we are done. Again, we prove by contradiction: Assume at time t that $\forall n \in N, D'_n(t) \leq D_n(t)$, then:

$$[D_A(t) - D'_A(t)] + \sum_{n \in N} [D_n(t) - D'_n(t)] > 0, \tag{4.9}$$

which contradicts Eq. (4.8). ■

Theorem 2 tells us that in order to be successful with a theft attack that evades the balance check, it is necessary for Mallory to ensure that the consumption of at least one of her neighbors is over-reported. She can achieve the equivalent of that by compromising a neighbor's smart meter or by physically tapping into the neighbor's electrical system. Both methods could ensure that the neighbor pays for Mallory's electricity. If the neighbor were generating electricity, Mallory would under-report the neighbor's generation. The notation still holds because that attack is equivalent to over-reporting of consumption.

Let L_n denote the monetary loss incurred by a neighbor n who has been targeted by this attack. Then,

$$L_n = \Delta t \sum_{t=1}^{T} \lambda(t)[D'_n(t) - D_n(t)]. \tag{4.10}$$

The structure of Eq. (4.8) shows that the amount of electricity Mallory can steal at time t is maximized when she steals as much electricity as she can from all her neighbors. If she compromises more than one neighbor, then the total amount of electricity she steals in T time periods is given by $\Delta t \sum_{n \in N} \sum_{t=1}^{T} [D'_n(t) - D_n(t)]$, and the total monetary worth of this electricity is given by α in Eq. (4.2) as $\alpha = \sum_{n \in N} L_n$. Note that if $D_n(t) = D'_n(t)$, then Mallory has not stolen electricity from neighbor n at time t.

If the balance check were in place, Attack Classes 1A, 2A, and 3A could be implemented only if Mallory performed the additional step of over-reporting a neighbor's consumption.

With that additional step, Attack Classes 1A, 2A, and 3A would be renamed to *Attack Classes 1B, 2B, and 3B*, where the letter "B" is used to indicate that they circumvent balance checks. Those attacks are summarized in Table 4.1, and illustrated later in Fig. 5.7. We now describe a fourth class that involves compromise of the electricity price in addition to consumption readings.

Attack Class 4B: This class of attacks illustrates how a neighbor can receive a lower electricity bill than expected despite having electricity stolen from him. This strategy can only work in a real-time pricing setting in which consumers are equipped with Automated Demand Response (ADR) interfaces. ADR encourages electricity consumers to adapt to changes in signals from the utility, most importantly the electricity price. This ensures that demand is adjusted to meet supply constraints. The most well-known implementation of ADR, known as OpenADR, is based on the Energy Market Information Exchange (EMIX) specification [87].

In Attack Class 4B, Mallory actively decreases her neighbors' demand by compromising their ADR interfaces. Here, she increases her consumption in proportion to the amount by which she decreases her neighbors' consumption. She effects the decrease of her neighbors' consumption by increasing the electricity price seen by the neighbors' ADR systems. Since consumption is typically modeled as a monotonically decreasing function of the price, an ADR system of a consumer would be programmed to automatically consume less if the electricity price has increased. The Consumer Own Elasticity model [88] is an example of such a monotonically decreasing function that captures how much a consumer would decrease his consumption in response to a given increase in the price.

Therefore, the attack is designed in such a way that for some time t and neighbor n, $D_n(t) < D'_n(t)$, $D_A(t) > D'_A(t)$ and $\lambda(t) < \lambda'_n(t)$, where $\lambda'_n(t)$ is the compromised electricity price seen by neighbor n. Based on his meter's readings, the unsuspecting n thinks he has consumed $D'_n(t)$ and expects to pay a bill of B_{Expected} during a billing cycle of T time periods. Instead, he is sent a lower electricity bill by the utility, B_{Utility}, which leads him to believe

that he benefited by a positive quantity ΔB:

$$\begin{aligned}\Delta B &= B_{\text{Expected}} - B_{\text{Utility}} \\ &= \Delta t \sum_{t=1}^{T} \lambda'_n(t) D'_n(t) - \Delta t \sum_{t=1}^{T} \lambda(t) D'_n(t) \\ &> 0.\end{aligned} \tag{4.11}$$

This is interesting since, in reality, he lost a positive amount to Mallory, L_n, given by Eq. (4.10). The notation holds in the case where this attack has been launched against multiple neighbors. Those neighbors who have not been compromised would have $D_n(t) = D'_n(t)$, and $\lambda(t) = \lambda'_n(t)$.

4.6 Framework for DER Fraud

The meter readings of a generator C represent the *average net generation* $G_C(t) \in \mathbb{R}$ during each time period t, and are in kW/MW. $G'_C(t)$ is the reported value corresponding to $G_C(t)$. If $G'_C(t) \neq G_C(t)$, the meters are not reporting their actual values and must be investigated.

Let $\lambda(t)$ denote the electricity price during the time period t, where $\lambda(t) \in \mathbb{R}$. Note that the price does not necessarily change between smart meter polling periods, and that price updates are usually less frequent than polling reports. We assume that for any time period t the electricity price $\lambda(t)$ is common to all customers.

The attackers' monetary advantage through fraud, α, is given by the difference between what the utility should pay them based on the actual generation, B_{Utility}, and what the utility actually pays them based on the reported generation, B'_{Utility}. If the billing cycle contains T time periods, then the attacker, A, can make a monetary gain through fraud if

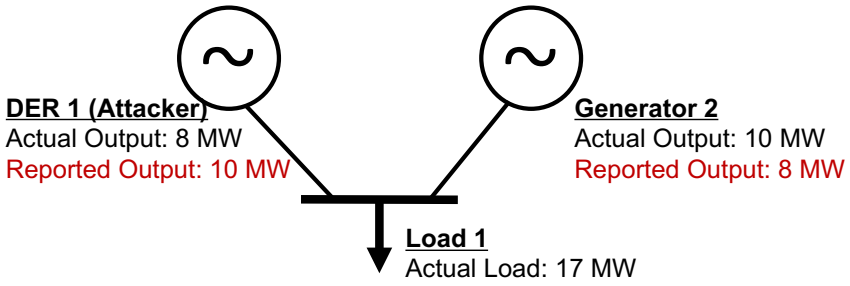


Figure 4.2: How attackers can circumvent the balance check by over-reporting their own generation and simultaneously under-reporting another generator's output (or by over-reporting the load).

and only if the following condition holds:

$$\begin{aligned}\alpha &\triangleq B'_{\text{Utility}} - B_{\text{Utility}} \\ &= \sum_{t=1}^{T} \lambda(t) G'_A(t) \Delta t - \sum_{t=1}^{T} \lambda(t) G_A(t) \Delta t \\ &> 0,\end{aligned} \tag{4.12}$$

where the units may be given as follows: λ is in \$/kWh, G is in kW, Δt is in hours, and α is in \$ (dollars). The attackers' objective is to maximize α subject to the constraint that the attack must go undetected. Since $\Delta t > 0$, Eq. (4.12) holds only if $\text{sgn}(\lambda(t))[G'_A(t) - G_A(t)] > 0$ for some t, where sgn is the sign function. The statement is evident and the proof follows from the proof of Theorem 1. Therefore, the attackers must over-report their generation in order to make a monetary gain when the price is positive and under-report when it is negative. We design attacks for the far more common case in which $\lambda(t) \geq 0$.

A naive way to detect such an attack would be to use a variant of the balance check, described in Section 4.4.1, that would use redundant meters to ensure that the total amount generated is the total amount consumed. The balance check is naive because it can easily be circumvented as follows. Consider the schematic diagram in Fig. 4.2. The two generators illustrated are logically separated. Each generator may be composed of multiple individual generators whose values sum up to the values shown in the figure; the same applies to loads. DER 1 describes a group of DERs whose reported output exceeds the actual output, while G

2 describes a group of generators whose reported output is less than the actual output. The load may also be misreported, but that is not illustrated in the figure for simplicity. Energy is conserved because the total actual generation is equal to the total reported generation. That value (18 MW) is the sum of the load (17 MW) and losses due to electric line impedance and transformer cores (1 MW).

4.7 Detection Model

In the context of this dissertation, a detection method is a centralized online algorithm that would run at an electric utility's control center. The detection methods discussed in this section look for anomalies in the smart meter readings that are reported to the utility and stored at the control center. We assume that the meters perform accurate measurements. This assumption is justified by a study [85] that concluded that 99.96% of electronic smart meter readings were within ±2% of the actual value, and that 99.91% were within ±0.5%. Therefore, an attacker cannot leverage measurement errors inherent to smart meters to steal a significant amount of electricity. *While the measurements are accurate, the reported readings may have been compromised by an attacker.*

If the smart meters can be compromised, it is reasonable to assume that the meters performing the balance check at the internal nodes of the network topology can also be compromised. We assume that the balance meter at the root node of the network alone can be trusted. This assumption is easily justified if this meter is located in a substation on the same premises as the utility-owned control center. The control center may be primarily tasked with substation automation, but it can also be used to detect electricity theft. Since the balance meter at the root node and the control center are co-located, the meter may directly feed into the servers at the control center by using dedicated communication infrastructure that is not exposed to external attacks.

Under the assumption that the root node balance meter is trusted, Attack Classes 1A–3A are automatically detectable using the balance check. Therefore, no further work needs to be done in order to detect these classes of attacks. However, Mallory can still use Attack Classes 1B–4B to steal electricity from her neighbors. *We focus on detecting Attack Classes*

1B, 2A, 2B, 3A, and 3B. While Attack Classes 1A–3A can be detected using the balance check, Attack Classes 2A and 3A can also be detected using the same data-driven methods that we use to detect Attack Classes 2B and 3B. *Attack Class 1A cannot be detected by data-driven methods since the reported consumption in this class of attacks is in no way abnormal*. Also, we have no data to support an ADR-based system in the case of Attack Class 4B. In order to study Attack Class 4B, we would need to make assumptions about how each consumer in the dataset changes consumption in response to changes in real-time electricity prices. We would also need to simulate a real-time electricity market, and that is beyond the scope of this work.

4.7.1 Nature of Anomalies due to Attack Injections

Abnormally high consumptions may indicate that the owner of the smart meter is a neighbor of the attacker in one of the Attack Classes 1B–3B. Under Attack Class 1B, Mallory reports normal consumption readings but consumes more electricity than reported. This extra electricity is billed to her neighbors, whose consumptions are over-reported. In order to identify Mallory, we would need to identify her neighbors and then manually validate all meters connected to their parent node. As mentioned earlier, Attack Class 1B is the most severe of all classes, as Mallory can steal an arbitrary amount of electricity from her neighbors. The only limit on how much she can consume is determined by the physical limits of the electrical conductors in the distribution lines that connect her facility to the grid.

Abnormally low consumptions, as would be reported under Attack Classes 2A and 2B, are a characteristic of the attacker, and would help us identify Mallory herself. Since consumption readings are under-reported in both Attack Classes 2A and 2B, we group the two classes together as 2A/2B and apply the same techniques to detect abnormally low consumption.

Attack Classes 3A and 3B both involve load shifting from a period of high prices to a period of low prices. We show that abnormal consumption *for a given price* indicates that an attack from either of these two classes is happening, and we group the classes together as 3A/3B for detector evaluation.

4.7.2 Detection Procedure

We have discussed three attack classes that fail balance checks and four that circumvent these checks. Mallory could use any combination of these attack classes in her actual attack, so detection methods cannot be designed to target individual classes. With F-DETA, we propose a holistic approach to detection that works on *all* the classes.

The five general steps for detecting all seven attack classes are (1) to use a model to estimate the expected consumption of all consumers for the upcoming time period; (2) to evaluate whether the actual readings obtained are anomalous based on what was expected; (3) to identify whether the anomalies indicate that the consumer is an attacker (abnormally low readings) or a victimized neighbor of an attacker (abnormally high readings) as per Theorem 1; (4) to use external evidence (severe weather conditions, holiday periods, special events, etc.) to determine whether the anomalous consumption may be a false positive; and (5) to investigate the anomaly systematically, if there is no reason to suspect a false positive, by checking the integrity of the smart meters as described in Section 4.4.2.

4.8 Related Work

In this section, we present prior and ongoing efforts by the research community and industry to detect and defend against electricity theft attacks. Electricity theft detection methods include those based on well-defined attack strategies [47, 86, 89] and general consumption behavior anomalies [78].

In [86], the authors evaluate a few different attack detection algorithms for a very specific electricity theft strategy in which the attacker does not change her consumption behavior, but reports lower consumption readings by compromising her own smart meter. In [47], we evaluated a different attack strategy wherein the attacker steals electricity from a neighbor at no loss to the utility. Those two papers failed to capture other possible attack classes because a comprehensive and fundamental approach to classification was not adopted. We fill in that gap with a framework that provides a comprehensive classification for better defense. In [90], [91], and [84], the authors assume that smart meters have not been compromised,

and use their readings to detect electricity theft. They do so by calculating the total power lost and estimating how much of the loss was due to electricity theft. Their methods fail under the realistic scenario in which smart meters have been compromised, and we address this gap. The motivations of attackers who steal electricity are discussed in detail in [91].

Industry has also invested in mitigating electricity theft. Utilities such as BC Hydro and CenterPoint have implemented tamper detection features on smart meters [92]. Unfortunately, penetration testing on a variety of different smart meters has shown that such features are ineffective [93], and that despite decades of work on tamper detection schemes (the first of which was the patent [94]), better protections against electricity theft are needed. BC Hydro has worked with startup Awesense to go one step further than tamper detection by placing distribution grid meters, which are different from consumer smart meters, at key nodes on BC Hydro's distribution grid [92]. Although those efforts have been tailored to address line-tapping electricity theft, we showed that the investment in distribution grid meters can also be effective against cyber intrusion-based theft attacks.

4.9 Conclusion

In this chapter, we developed a comprehensive theoretical framework to provide a defender with an understanding of possibilities for electricity theft attacks on smart grids. We applied that framework to a large-scale smart-meter dataset to devise data-driven detection mechanisms that mitigate those attacks. In doing so, we partially addressed the second research objective stated in Chapter 1.4 by providing actionable insights for improving fraud detection in smart grids. We will continue to explore fraud detection and completely address that research objective in the following three chapters. The work in this chapter was peer-reviewed and published in [95].

CHAPTER 5

DATA-DRIVEN DETECTION AND MITIGATION OF CONSUMPTION FRAUD

In this chapter, we use the framework developed in Chapter 4 to detect consumer fraud in AMI. In particular, we validate the data reported to the utility by modeling the normal consumption patterns of consumers, and detecting deviations from this model. Our models are data-driven, and use readings from a real smart meter deployment in Ireland.

It must be noted that smart meters do come with security features, which make them difficult for an attacker to compromise. For example, meters manufactured by GE [96] are equipped with encrypted communication capabilities and tamper-detection features. However, reliance on those mechanisms alone is not sufficient to ensure total defense against cyber intrusions that exploit communication vulnerabilities. Methods to circumvent encrypted communications in smart grid protocols were recently presented in [97]. We assume that the attacker, whom we continue to refer to as Mallory, can exploit one or more existing vulnerabilities in AMI to commit meter fraud. The logistics of how she can get into a position where she is capable of modifying communication signals is not a focus of this chapter and is discussed in [78], [97], [98], [99], and [93].

Before exploring fraud detection, let us briefly describe how fraud may be prevented. The aforementioned vulnerabilities in smart meters could be patched to prevent Mallory from gaining access to smart meters and their communication messages. However, that is impractical because hundreds of millions of smart meters have been deployed worldwide and they all have those vulnerabilities. Furthermore, there is no secure mechanism in place to automate the installation of patches over communication networks. Therefore, the patching would have to be a manual procedure. Even if the patches were somehow installed, the aforementioned vulnerabilities are only the known vulnerabilities, and there may be many unknown vulnerabilities that Mallory may exploit. The unknown vulnerabilities can only be

patched once they are discovered, but until then they can be exploited by Mallory. Our aim goal in this chapter is to detect attacks under the conservative assumption that Mallory has successfully managed to compromise the integrity of smart meter consumption readings.

5.1 Summary of Contributions

Using F-DETA, proposed in Chapter 4, we evaluate fraud detection methods proposed in the literature and propose four new fraud detection methods in this chapter. We use a dataset of smart meter readings from Ireland to demonstrate the detection methods, and that dataset is described in Section 5.2. Before describing the four detection approaches, we provide a broad characterization of approaches for detecting anomalies in time-series data in Section 5.3. Fraud detection is accomplished entirely through the detection of anomalies in power consumption data; we had access to no other kinds of data in our study.

Before proposing our own detectors, we discuss a detector proposed in related work in Section 5.4. Two of the fraud detectors that we proposed in this chapter are based on the autoregressive integrated moving average (ARIMA) model. They are described in Section 5.5. The third detector, described in Section 5.6, is based on a novel combination of principal component analysis (PCA) and density-based clustering of applications with noise (DBSCAN). The fourth detector, described in Section 5.7, uses Kullback-Leibler (KL) divergence to measure the difference between two distributions of meter readings for anomaly detection. We evaluate the PCA-DBSCAN and KL divergence detector against the average detector in Section 5.8 and against the ARIMA-based detectors in Appendix B. We present related work in Section 5.9 and conclude the chapter in Section 5.10.

5.2 Description of the Dataset

The dataset we used was collected by Ireland's Commission for Energy Regulation (CER) as part of a trial that aimed to study smart meter communication technologies. It is the largest publicly available dataset that we know of. The fact that the dataset is public makes it possible for researchers to replicate and extend our results. The dataset is an anonymized

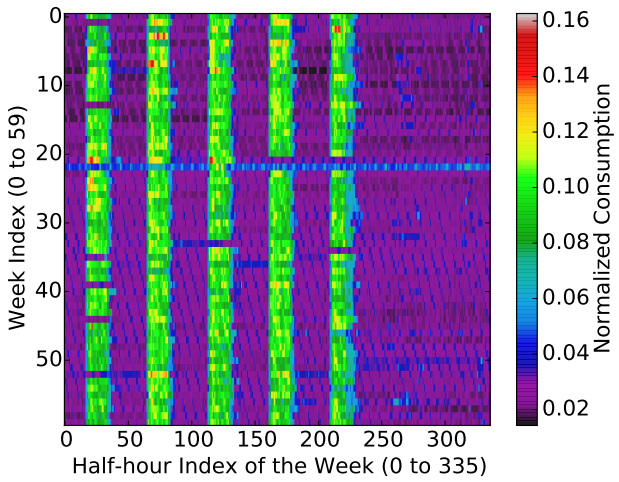


Figure 5.1: Normalized consumption of a commercial consumer. The five green/blue vertical bands represent time periods of higher electricity consumption, and they correspond to weekday business hours.

collection of smart meter readings from consumers, collected at a half-hour time resolution, for a period of up to 74 weeks. Our set of 500 consumers includes 404 residential consumers, 36 small and medium enterprises (SMEs), and 60 unclassified by CER.

We assume that the dataset obtained from CER is free from integrity attacks. However, there are anomalous consumption behaviors in the dataset. These behaviors might reflect periods when consumers were traveling (leading to abnormally low consumption), hosting parties (leading to abnormally high consumption), or engaging in other unusual but legitimate activities. Such events lead to false positives if the detection strategy classifies them as suspected attacks.

We divided the 74 weeks of consumption data obtained from the CER dataset into two sets: a *training set* of the first 60 weeks, and a *test set* of the remaining 14 weeks. Figure 5.1 illustrates the training set corresponding to the consumption of one consumer in the dataset. Each row corresponds to one week of data, so there are 60 rows corresponding to the 60 weeks in the training set. The rows are normalized by their $l-2$ norms, and the colors represent the normalized consumptions over the week. There are 336 columns, each of which corresponds to one half-hour period of the week.

Note that anomalies in the training set are not labeled, so we do not have ground truth on which readings are anomalous. Thus, our algorithm is essentially unsupervised, and

our training set serves to build a model of the consumption patterns while accounting for the possibility of anomalies in it. For example, the row in Fig. 5.1 corresponding to week index 23 is clearly anomalous; it corresponded to the Christmas-New Year period, when consumptions can be expected to deviate for the holiday season. Such weeks produce false positives in the dataset, and the test set is used to evaluate the false positive rate.

As the dataset does not contain the electric distribution network topology, we do not know which consumers share a parent node in the topology. As a result, the placement of balance meters is unknown, so we make the conservative assumption that *the balance meter at the root node is the only balance meter that has been deployed in the electric distribution network*. Irrespective of the network topology, all the consumers must be ultimately connected to the root node, so the sum of the readings from all consumers can be used to enforce the balance check at the root node. Equivalently, we could make the more complex assumption that *Mallory has compromised all balance meters in the topology except the one at the root node*. We justified this assumption in Section 4.7.

5.3 Detecting Anomalies in Time-Series Data

Before describing the proposed anomaly detection methods for detecting consumption fraud, we characterize anomaly detection methods for time-series data in general. Broadly speaking, there are two approaches one can take for performing anomaly detection on time series data. First, an individual reading can be evaluated based on whether or not it is anomalous. Second, a set of readings as a whole can be evaluated on whether or not they are anomalous. We will explain both approaches next.

5.3.1 Detecting Anomalies in Individual Readings

We know from *Proposition 1* that, in order to launch a successful theft attack, $\exists t$ such that $D'_A(t) < D_A(t)$, and we know from *Proposition 2* that $\exists(n,t)$ such that $D'_n(t) > D_n(t)$ for $n \in N$. So at each time period t, we need to check how far the smart meter reading $D'_c(t)$ is from our expected value of $D_c(t)$ $\forall c \in C$. The larger the difference, the larger the amount

of electricity that can be stolen. The ARIMA-based detection approach proposed later in this chapter detects anomalies in individual readings.

5.3.2 Detecting Anomalies in Sets of Readings

Anomalous behavior is apparent only when one observes a set of multiple consumption readings that deviate from the historic trends. In this chapter, we standardize the size of the set to a full week of 336 half-hour readings. That is a good choice of size, as consumers' weekly consumption patterns tend to repeat. Daily trends also exist; however, they are not as pronounced as weekly trends because of differences in schedules between weekdays and weekends.

It may appear that the use of multiple readings has a limitation in that an entire week of data needs to be collected before an anomaly can be detected. There are two counter-arguments to that point. The first is technical: the new week's vector can be completed with trusted data from a week in the training set (historic readings). As new consumption readings are recorded, they will replace the historic readings in the week's vector. If the week's vector contains sufficiently anomalous readings right at the beginning, it may appear anomalous before a full week of new data has been collected. The second counter-argument is based on litigation proceedings: if an electricity thief has been caught, the fines imposed may exceed the value of electricity stolen in a single week. Under that assumption, the week-long upper bound on the time-to-detection may be acceptable.

The detection approach proposed later in this chapter, which augments ARIMA-based detection with checks on mean and variance, performs anomaly detection on a set of readings. The mean and variance are calculated on that set. Similarly, the PCA-DBSCAN detector and the KLD detector also perform anomaly detection on sets of readings

5.4 Detection Using Averages

In this section, we describe the *min-average* detector proposed by Mashima and Cardenas in [86]. This detector finds a set of readings anomalous if the average of the set is less than the

minimum of the averages of sets of past readings. For example, if we were to take 60 weeks of consumption readings for a training set and focus our attention on a single consumer, then the first step would be to compute the averages of the consumption readings in each of those 60 weeks. Then, we compute the minimum of those averages and use *min_avg* to denote that minimum. That is the end of the training period. In the testing period, for a week of consumption readings in the test set (which may or may not be malicious), we compute the average and compare it with *min_avg*. If that average is less than *min_avg*, we mark that as being anomalous.

5.4.1 Min-Average Attack

Let us now derive the worst-case attack against the min-average detector; this is new work, and was not done by the authors of [86]. The worst-case attack for the detector is the optimal attack for the attacker (denoted by A). The optimal attack vector for a week of readings, denoted by D'^{*}_{A}, maximizes the attacker's profit as follows.

$$D'^{*}_{A} = \arg\max_{D'_A} \sum_{t=1}^{T} \lambda(t)\Delta t[D_A(t) - D'_A(t)] \tag{5.1}$$

$$= \arg\min_{D'_A} \Delta t \sum_{t=1}^{T} \lambda(t) D'_A(t) \tag{5.2}$$

$$\text{subject to} \tag{5.3}$$

$$D'_A \geq 0, \tag{5.4}$$

$$D'_A \leq max(D'_{Tr}), \tag{5.5}$$

$$\frac{1}{T}\sum_{t=1}^{T} D'_A(t) \geq \text{min\_avg}, \tag{5.6}$$

$$\tag{5.7}$$

where $\lambda(t)$ is the electricity price at time t and Δt is the time period between the collections of readings. The objective function in the first line is the profit given in dollars ($); Δt is in hours (h); $D_A(t)$ is in kilowatts (kW); and $\lambda(t)$ is in $/kWh. The equivalent objective

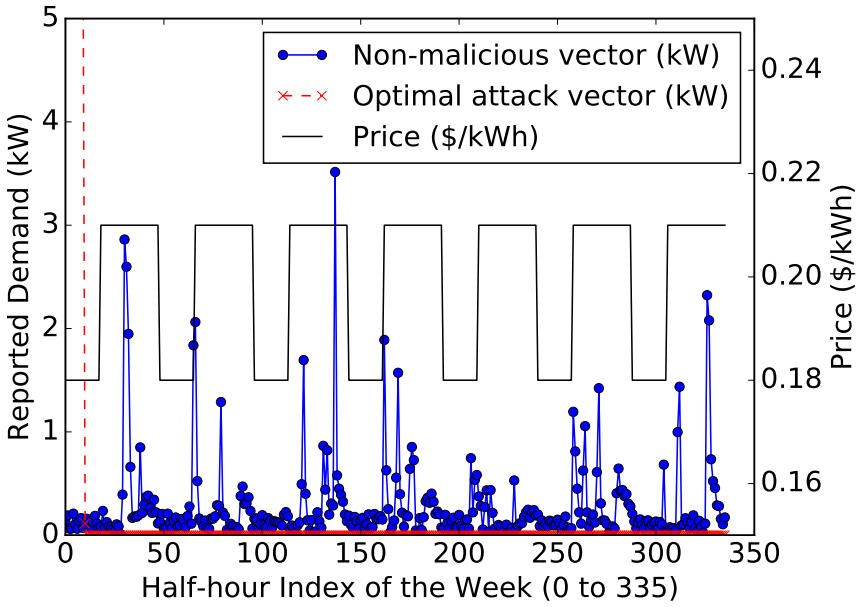


Figure 5.2: Illustration of optimal attack against PCA in the original dimension.

function in the second line denotes minimizing the attacker's utility bill. The first constraint is a non-negativity constraint on the reported readings D'_A. The second constraint ensures that the readings are less than the maximum of the values in the attacker's training set D'_{Tr}. The last constraint is the min-average detector constraint. Since the objective function and the constraints are all linear in D'_A, the optimization problem is a linear program and can be efficiently solved using a variety of freely available solvers.

The solution to the above optimization problem produces the worst-case attack against the min-average detector, and we refer to the attack as the *min-average attack*. The solution, D'^*_A, comprises a full week of readings, and is illustrated for one sample consumer in Fig. 5.2. Most of the readings were set to zero; the first few readings were set to $max(D'_{Tr})$ to ensure that the overall average was equal to *min_avg*. Also notice that the large values equal to $max(D'_{Tr})$ were injected at a time when the electricity price was low.

5.4.2 Impact of the Min-Average Attack

We evaluated the min-average attack by allowing Mallory to take the roles of the 500 consumers in the CER dataset one at a time and then averaging the results across all consumers.

We accomplished that by injecting the min-average attack separately for each consumer in a test set of 14 weeks of data, and computing the average gains for each consumer across those 14 weeks. The gains are defined by the difference between what the consumer would have paid as electric utility bills had she not committed fraud, and what she actually paid having committed fraud.

The result was that \$26.4 per week could be gained by an attacker (lost by the utility) due to fraud, on average across the 500 consumers. That is over \$1300 per consumer per year on average. In a country of 100 million consumers, of which 1% commit meter fraud, the losses to the utility would amount to over \$1.3 billion per year. Therefore, the average detector is not a strong enough deterrent for fraud, and we will propose detectors that *build on top of the min-average detector* to mitigate the min-average attack next.

5.5 Detection Using ARIMA Models

We propose to detect single-reading anomalies by using the Auto-Regressive Integrated Moving Average (ARIMA) model. In that method, historic data (from the training set) is used to forecast the next consumption reading in the time series $D_c(t)$. We proposed the ARIMA-based detector as a first-level check on the range of smart meter readings using the confidence interval obtained from the ARIMA model.

Auto-Regressive Moving Average (ARMA) models are simpler versions of ARIMA models, and their usefulness in detecting meter fraud was first proposed by the authors of [86]. We show that for most consumers, ARMA models are not suitable. The underlying assumption of the ARMA model is that the time series data is weakly stationary. Stationary data have three characteristics: (1) the mean is constant, (2) the variance is constant, and (3) the covariance of the signal with itself at different time lags is constant. We define a weakly stationary signal as one that fails condition (1), but satisfies conditions (2) and (3). The moving average component of ARMA automatically adjusts for changing means, so condition (1) is not important for the suitability of ARMA for a given time series.

For the electricity consumption time series of a single consumer at time t, given by the

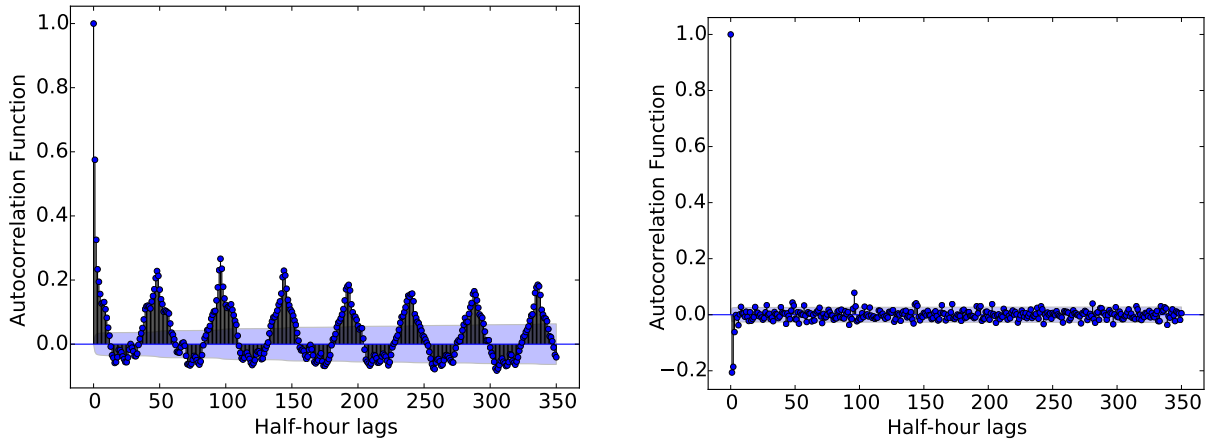
value of X_t, the ARMA model is defined as follows:

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \epsilon_{t-j}. \tag{5.8}$$

In the auto-regressive component, X_t is an affine function with intercept c of the time signal at p previous time points X_{t-i} with linear coefficients α_i. The moving average component of ARMA handles weakly stationary signals that do not have constant means. It assumes that i.i.d. Gaussian noise $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ compounds over q time periods to contribute linearly to the signal X_t with coefficients β_j.

The ARMA model does not handle largely changing covariance in non-stationary signals. Figure 5.3(a) illustrates the auto-correlation function (ACF) for a single consumer. The ACF is the correlation of the time series with itself at a specified lag. We extract the time series for a single consumer and depict the ACFs for 350 half-hour lags. There are 336 half-hours in a week, so the figure captures a little over a week. As expected, high auto-correlation was observed for this consumer at multiples of 48 half-hour (or 1-day) time periods. These high correlations persist for all lags throughout the consumption history captured in the dataset. Furthermore, the plot demonstrates failure of the third requirement for stationarity since the ACFs change significantly over time. This lack of stationarity implies that the ARMA model would fail to provide a reliable prediction of the next point in the time series. The ACFs need to rapidly decrease to constant or insignificant values in order for the ARMA model to reliably work. The rate of ACF decrease will determine the model order.

We propose an alternative model, the ARIMA model, which has an additional differencing term. We find that first-order differencing causes rapidly decreasing ACFs for consumers who have non-stationary consumptions. First-order differencing modifies the ARMA model in Eq. (5.8) as follows. Instead of predicting the next value in the time series, we predict the difference between the current and next values in the time series as a linear function of past differences.



(a) ACFs without differencing

(b) ACFs with first-order differencing

Figure 5.3: Autocorrelation function of the time series signal of a single consumer. The lag is in terms of half-hour time periods.

$$X_t - X_{t-1} = c + (\epsilon_t - \epsilon_{t-1}) + \sum_{i=1}^{p} \alpha_i (X_{t-i} - X_{t-i-1}) + \sum_{j=1}^{q} \beta_j (\epsilon_{t-j} - \epsilon_{t-j-1}). \tag{5.9}$$

In essence, a linear model fits the gradients of the points as opposed to the points themselves. After applying first-order differencing, we observe Fig. 5.3(b). Clearly, the ACFs are close to zero beyond 3 time lags. Therefore, the order of the ARIMA model is finite. In addition, the order is small (p and q are around 3), which is important to ensure minimal computational costs.

We have applied first-order differencing and observed its benefits for one consumer, but visual inspection is impractical for our dataset of 450 consumers. Therefore, we employ the Hyndman-Khandakar algorithm [100] to estimate the model order. This method combines cross-validation techniques, unit root tests, and maximum likelihood estimation.

The results of the Hyndman-Khandakar algorithm are as follows. While the autoregressive (p) and moving average order (q) range from 0 to 5, the differencing order is either 0 or 1. A minority of consumers (35 out of 450, or 7.78%) have stationary consumption patterns and thus the ARMA model proposed in [86] is appropriate for this minority. However, for 92.22% of consumers, first-order differencing is required, justifying our ARIMA model proposal. The distribution of consumers, segregated by consumer type, is captured in Fig. 5.4.

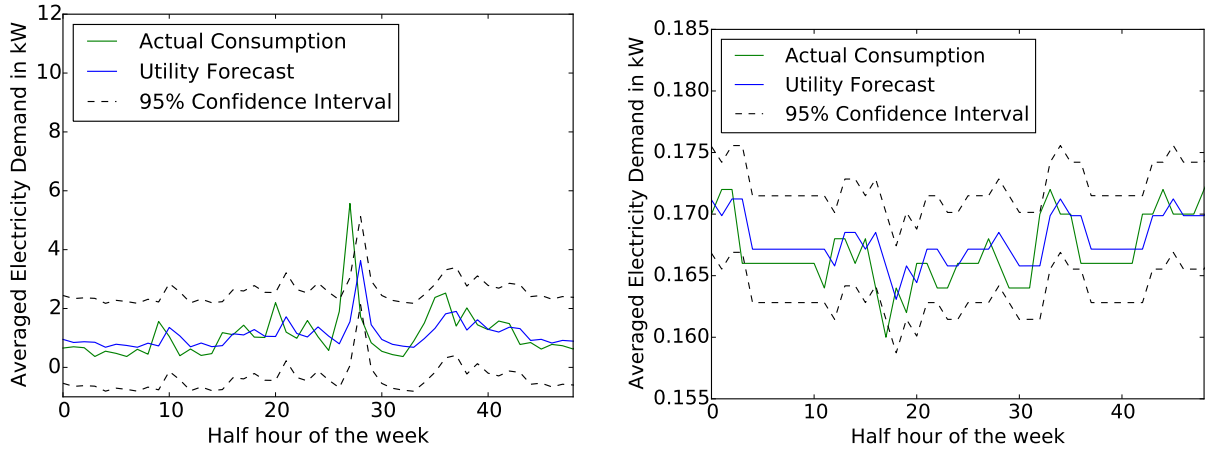
Figure 5.4: Distribution of differencing order among consumers of different types.

Once the ARIMA model is estimated, the next consumption point in the time series X_t is forecast. From that point forecast, a 95% confidence interval C is constructed with the assumption of i.i.d. Gaussian errors ϵ_t as described in [101]:

$$C = X_t \pm 1.96\sigma_\epsilon. \tag{5.10}$$

Here 1.96 comes from the fact that 95% of the standard normal distribution lies within [-1.96,+1.96]. Recall that σ_ϵ was the standard deviation of the i.i.d. Gaussian errors ϵ_t in Eq. (5.8). The prediction by ARIMA and the confidence intervals for two different consumers are illustrated in Fig. 5.5. In this chapter, we propose the use of these confidence intervals for anomaly detection. If a smart meter reading lies outside these intervals, we say with 95% confidence that it is anomalous. Also note that there is an order of magnitude difference between the consumptions of these two consumers and that the confidence intervals for Consumer 2 in Fig. 5.5(b) are tighter because of lower variance in consumption patterns. Tighter confidence intervals are preferred, as they make attacks easier to detect. We refer to the detector based on ARIMA confidence intervals as the *ARIMA detector*.

In our dataset, the consumers do not produce electricity and sell it back to the grid,



(a) ARIMA prediction for Consumer 1

(b) ARIMA prediction for Consumer 2

Figure 5.5: ARIMA forecasting of points and 95% confidence intervals.

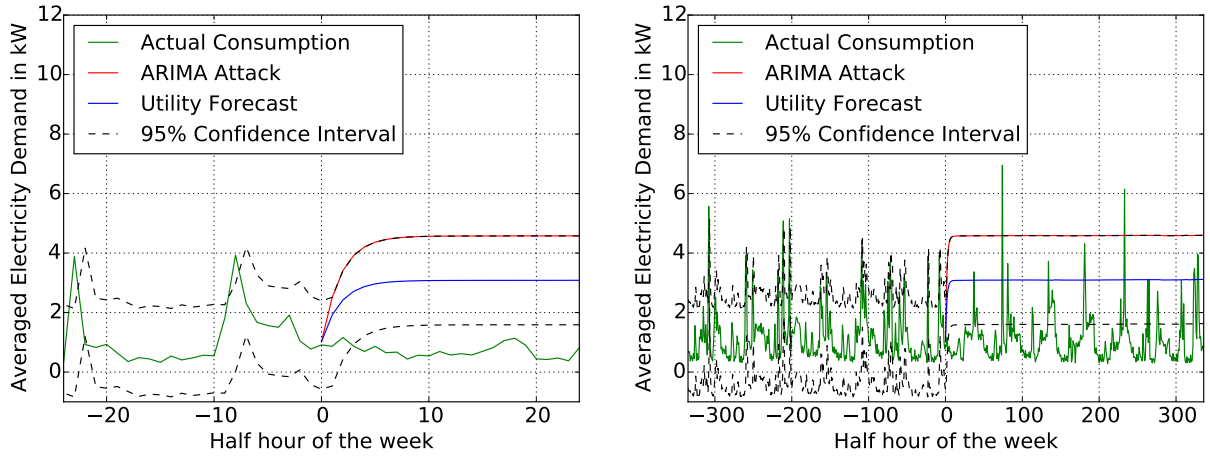
so the consumption is never negative. Thus, the lower bound of the confidence interval is useful only if it is positive, as negative readings reported by smart meters are naturally anomalous. Note that the lower bound for Consumer 1 in Fig. 5.5(a) goes negative, while it stays positive for Consumer 2. The reason why the lower bound goes negative is the symmetry in the Gaussian error assumption that is inherent in ARIMA and ARMA models. However, in future scenarios in which consumers supply to the grid, or consume a negative amount of electricity, a negative lower bound of the confidence interval will become useful.

5.5.1 ARIMA Attack

In order to circumvent the ARIMA detector, Mallory needs to ensure that their neighbor's consumption remains within the 95% confidence interval. If she steals more electricity from the neighbor, the utility will find that the neighbor's consumption exceeds the upper bound of the confidence interval and is anomalous. On discovering this anomaly, the utility may dispatch a technician to manually verify the integrity of the neighbor's meter. In practice, such investigations are made periodically [85]. In this section, we assume the worst-case scenario in which Mallory has full information and can estimate the 95% confidence intervals just as well as the utility can.

The optimal value for Mallory to steal is the maximum that she can steal while averting

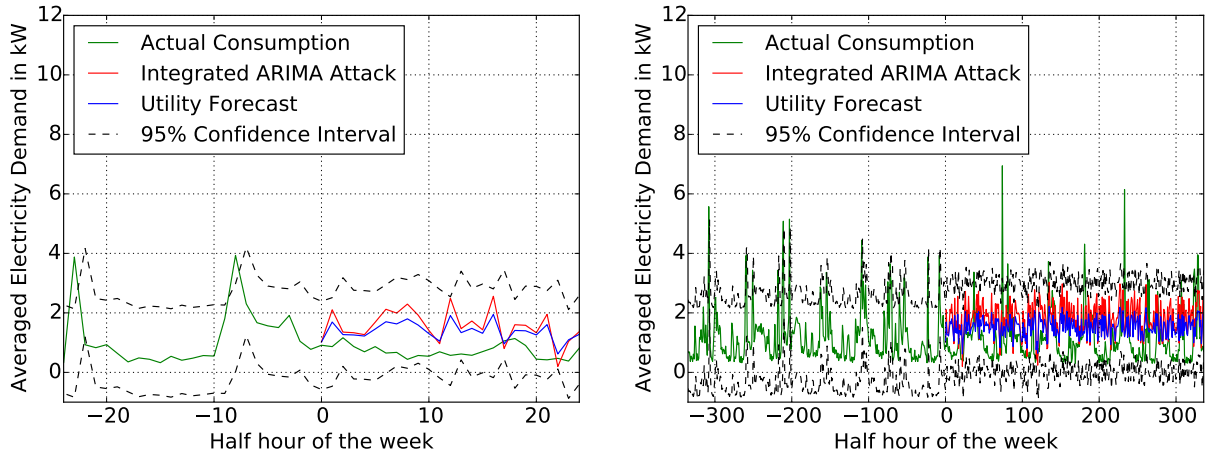
detection. This point is reached at the 95% confidence threshold. Thus the attacker over-reports the neighbor's consumption as the 95% threshold point as shown in Fig 5.6. Since this attacks averts the ARIMA detector, we refer to it as the *ARIMA attack*.



(a) Zoomed in (±12 hours)

(b) Zoomed out (±1 week)

Figure 5.6: Illustration of an ARIMA attack on a neighbor. The attack is launched at time 0 on the horizontal axis.



(a) Zoomed in (±12 hours)

(b) Zoomed out (±1 week)

Figure 5.7: Illustration of integrated ARIMA attack on a neighbor using the truncated normal distribution. The attack is launched at time 0 on the horizontal axis.

The ARIMA detector has bounded the attack, and the maximum electricity stolen from the neighbor is given by the difference between the ARIMA Attack curve and the Actual Consumption curve.

In order to detect the attack, the statistics of the window can be compared against

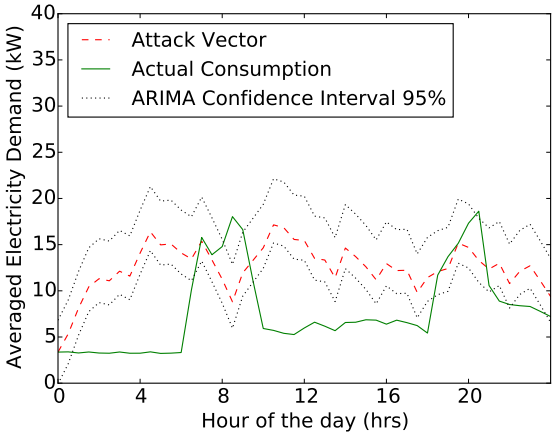
statistics of previous windows. Specifically, if the observed mean μ' lies in the interval $[min(\{\mu\}), max(\{\mu\})]$ and the observed standard deviation σ' lies in the interval $[min(\{\sigma\}), max(\{\sigma\})]$, then we say the new point is not a suspected attack. When the ARIMA detector is augmented with these additional checks on mean and standard deviation, we refer to the enhanced detector as the *integrated ARIMA detector*. Here $\{\mu\}$ and $\{\sigma\}$ are the sets of means and standard deviations observed in historic data. For the sake of standardization, we assume in our simulations that each statistic (μ and σ) is calculated on a window of a week in the historic data. Therefore, the cardinality of $\{\mu\}$ and $\{\sigma\}$ is the number of weeks in the utility's smart meter data archive.

5.5.2 Integrated ARIMA Attack

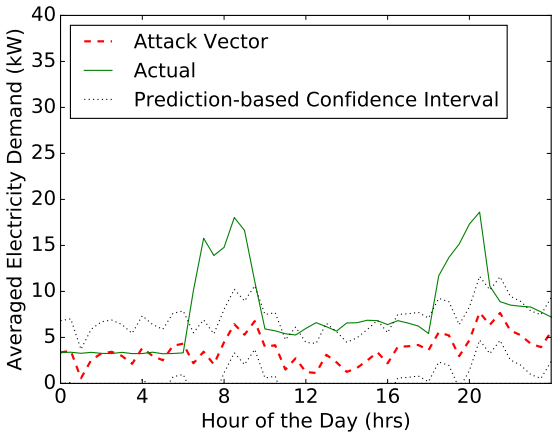
As security researchers and practitioners, it is important for us to think about how an attacker may evade our own checks, as no check is completely foolproof. In our case, we find that despite checks on the mean and standard deviation, it is possible for attackers to circumvent the integrated ARIMA detector. They may do so by generating false consumption readings by using a truncated normal distribution. This distribution is specified by a range, mean, and standard deviation. By setting the range to the ARIMA confidence intervals, Mallory averts detection by the ARIMA detector. In addition, she can set the mean to the extreme point $max(\{\mu\})$ to avert the check on the mean. At the $max(\{\mu\})$ value, the mean quantity of electricity stolen is maximized, while still being undetectable. Finally, Mallory can set the standard deviation to the extreme point $min(\{\sigma\})$ to avert the standard deviation check. We assume that she wants to minimize the standard deviation to minimize the variability that she needs to incorporate into his own consumption. If Mallory were to steal electricity from multiple consumers, the variability would add up, making it difficult for her to match with her own consumption in order to pass the balance check.

Since this attack averts all the checks of the integrated ARIMA detector, we called it the *integrated ARIMA attack*, and it is illustrated in Fig. 5.7. It can be seen that the integrated ARIMA attack curve has significantly higher variance than the ARIMA attack curve in Fig. 5.6. However, its mean is lower, so we expect less electricity to be stolen under the

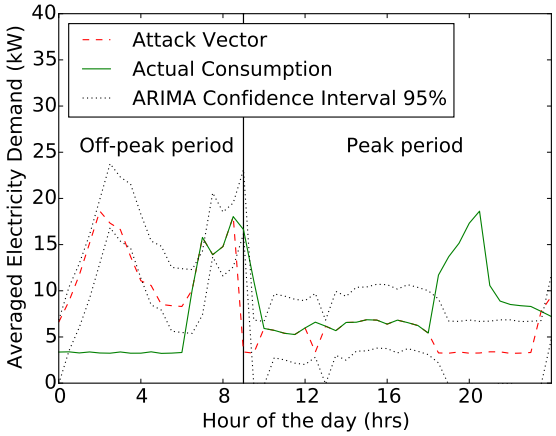
integrated ARIMA attack. The trade-off for Mallory is that the integrated ARIMA attack is harder to detect.



(a) Attack Class 1B (subject is a neighbor)



(b) Attack Classes 2A/2B (subject is the attacker)



(c) Attack Classes 3A/3B (subject is the attacker)

Figure 5.8: Illustration of Attack Classes 1B and 2A/2B using the *Integrated ARIMA attack*, and Attack Classes 3A/3B using the *Optimal swap attack*. In (a) the consumption of one of Mallory's neighbors is over-reported; in (b) Mallory's own consumption is under-reported; and in (c) Mallory's own highest consumptions are swapped into the off-peak period.

The various attack classes discussed in Chapter 4 can be realized as attack vectors by Mallory in many ways as long as she obeys the class definitions. For example, Mallory can under-report her consumption readings in Attack Classes 2A/2B by setting all reported readings to zero. Thus, she maximizes the amount of electricity that she can steal. However, it is easy to detect such an attack, and we want to inject attacks that are not as easy to

detect. We inject attacks using random numbers, as this ensures that deterministic patterns do not emerge, leading to easy detection. Note that the reported attack consumption poisons the utility's ARIMA model, so the confidence intervals follow the attack vector, and not the actual consumption. Figure 5.8 illustrates the injections through the use of the integrated ARIMA attack, where Mallory is Consumer 1330 in the CER dataset.

The injections were performed on the 500 consumers in the dataset. For each consumer, we injected attack vectors according to methods that we will describe next. For the integrated ARIMA attack, we generated 30 to 50 different attack vectors from the truncated normal distribution, to reduce bias in the samples obtained from the distribution. From these attack vectors, we evaluated all the detectors by using the worst-case vector that provided Mallory with maximum profit. The computation that went into producing this chapter required that 74 CPU cores be run for a total period of 4 weeks.

5.5.3 Attack Class 1B Injection

If we assume that Mallory can compromise a smart meter, it is also reasonable to assume that she can passively monitor it and build the same models of the data that we have built from the training set. If we were to install the ARIMA-based detector defined in Chapter 5, we could assume that Mallory can do the same.

The integrated ARIMA attack involves the injection of false readings from a truncated normal distribution in such a way that the neighbor's readings are over-reported, while remaining within the ARIMA confidence interval. At the same time, the mean and variance of the false readings do not exceed thresholds based on historic data. This attack is illustrated in Fig. 5.8(a).

5.5.4 Attack Classes 2A/2B Injection

We show that both the ARIMA attack and the integrated ARIMA attack can be implemented to realize Attack Classes 2A/2B. Mallory under-reports her own consumption readings to reduce her bill. For Attack Class 2B, she also over-reports consumption readings for her

neighbors to circumvent the balance check. In the case of the ARIMA attack, Mallory's attack vector would be set to the lower confidence threshold (or zero, whichever is greater). The ARIMA attack can be detected by the integrated ARIMA detector, which in turn can be circumvented by the integrated ARIMA attack as described in [47]. There is only one difference in injecting the integrated ARIMA attack for Attack Classes 2A/2B, and that is that the mean of the attack vector generated by the truncated normal distribution is equal to the minimum of the means (as opposed to the maximum, as for Attack Class 1B) in the training set. This injection attack is illustrated in Fig. 5.8(b).

5.5.5 Attack Classes 3A/3B Injection

In Attack Classes 3A/3B, Mallory reportedly swaps her load to exploit variable pricing. Injecting an attack vector requires us to assume either time-of-use (TOU) or real-time pricing. We assume TOU as there is currently no real-time pricing plan for end-consumers in Ireland (which is where our data came from). Based on Nightsaver plans offered by various Irish utilities, we assume two TOU periods: a peak period from 9:00 A.M. to midnight and an off-peak period from midnight to 9:00 A.M. We evaluated this choice of peak period for our dataset and found it to be suitable because 94.4% of consumers had higher consumption during the peak period on over 90% of the days in the training set.

We injected an optimal instance of Attack Classes 3A/3B that we call the *Optimal Swap attack*. We took a week of readings from the test set and swapped the highest readings from the peak period with the lowest readings in the off-peak period. Note that this attack does not affect the mean or variance of the data for the day (or for the week). It does not even affect the distribution of readings, so the KLD detector would not work if it were designed to compare the new week's vector with previous weeks' vectors. *The only change in the dataset is the temporal ordering of readings.* That ordering exploits the changing electricity price, as Mallory would pay less for her highest consumption readings in the day. The Optimal Swap attack would, however, require Mallory to be able to perfectly predict future high consumption readings so she can swap them with current low consumption readings in the early, off-peak period of the day. An imperfect prediction could also produce the same result,

but the prediction would need to be good enough to ensure that the distribution of readings was not significantly changed. For our study, we used prior knowledge of the values in the test set to make a perfect prediction, and injected the Optimal Swap, or worst-case scenario. Note that we injected the swapping of small values for large ones in a way that minimized errors that resulted from exceeding the confidence intervals of the ARIMA detector. That attack injection is illustrated in Fig. 5.8(c).

5.6 Detection Using PCA and DBSCAN

From Fig. 5.1, it is clear that the weeks of consumption readings look so similar that the matrix representation of consumption readings has a clear low-rank approximation. In other words, there exists a matrix with linearly dependent columns that looks almost identical to the one illustrated in Fig. 5.1. The difference is that the matrix illustrated in the figure contains deviations from the low-rank behavior as a consequence of small variations between weeks (noise). We propose to use principal component analysis (PCA) to discard that noise, and observe the similarity between weeks in a lower-dimensional space.

PCA reveals the underlying trends in the smart meter data, across thousands of consumers, by reducing the dimensionality of the data, while retaining most of the data's variance. As such, it provides us with a way to project a vector of electricity consumption readings in a high-dimensional space into one in a lower-dimensional space without loss of important signal characteristics. That greatly aids anomaly detection methods, which can be intuitively executed in the lower-dimensional space because clusters that differentiate normal consumption from abnormal consumption are evident in the lower-dimensional space.

5.6.1 The PCA Mechanism

We demonstrate the mechanism of PCA by constructing two different matrices (A & B) from our entire training set. A has $M_A = 20,160$ rows, one for each half-hour of the 60-week period of study, and $N_A = 2,982$ columns, one for each consumer. In this example, we can think of the consumption of each consumer across all 60 weeks as a column vector

in a 20,160-dimensional space. There are 2,982 such column vectors. Using PCA, we will collapse these column vectors from $M_A = 20,160$ dimensions into two dimensions. Because of high correlation, two data points are sufficient to capture the patterns of each consumer, relative to the patterns of other consumers. Let P_A be the matrix of dimension $2 \times M_A$ that transforms A of dimension $M_A \times N_A$ to Y_A of dimension $2 \times N_A$. Then,

$$P_A A = Y_A. \tag{5.11}$$

For calculation and notation convenience, we pre-process A in two stages. First, we center A by subtracting the mean of each row from all values in that row. Next, we divide A by $N_A = 2,982$. We are interested in the covariance between the M_A rows (or readings per consumer) in A. For the corresponding AA^T covariance matrix, $P_A = U^T_{(0:1,:)}$, where U is obtained from the singular-value decomposition (SVD) of $A = U\Sigma V^T$. Here, the columns of U are the eigenvectors of the covariance matrix AA^T, and Σ^2 (the eigenvalues) represent the amount of variance retained by the principal components. This is illustrated in Fig. 5.9. Together, the two components in P_A retain 63.63% of the variance in A, and the marginal variance retained by each additional component is negligible.

There are two advantages to retaining only the first two components. First, maximum variance is retained by these components, so discarding additional components effectively discards the noise in the consumption patterns. Second, it allows us to visualize a vector of $20,160$ dimensions in a two-dimensional space. That then facilitates anomaly detection in that 2D space, as we will discuss later. Later, we will evaluate different projections in more than two dimensions.

5.6.2 Construction of PCA Biplots

We transform A into the P_A space by taking the product $P_A A = Y_A$, where Y_A is the $2 \times N_A$ PCA score matrix. The two rows of Y_A are called the *Principal Component 1 Score* and the *Principal Component 2 Score*. The scatter plot of the two scores is the PCA biplot shown in Fig. 5.10(a). Points that lie close together in this 2D space describe consumers, or columns

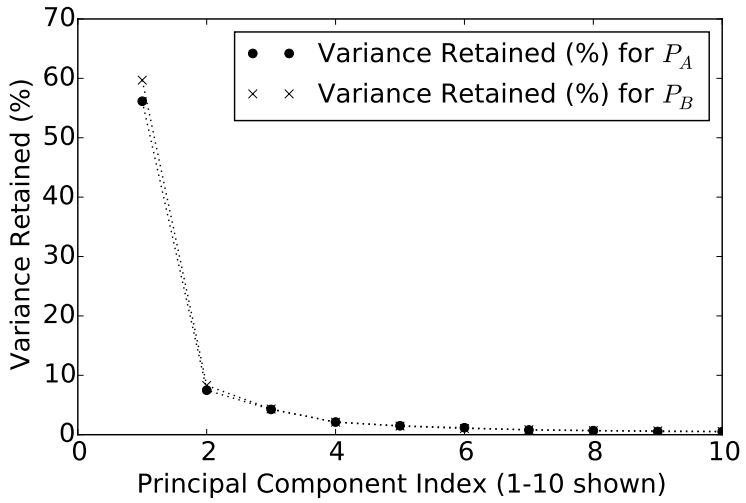
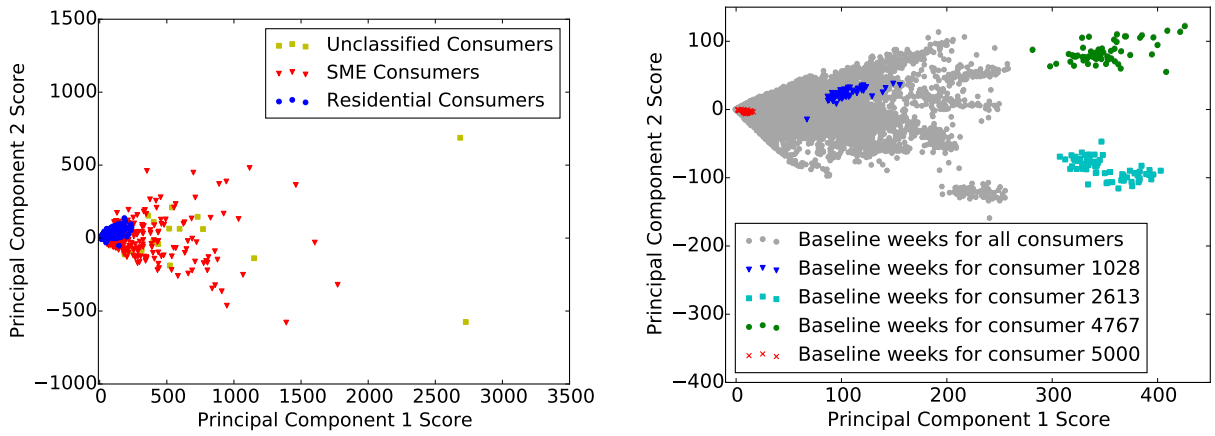


Figure 5.9: Variance (%) retained by principal components of matrices A & B.



(a) Observed clustering of consumers by type in the P_A space

(b) Observed clustering of consumers and their consumption weeks in the P_B space

Figure 5.10: Principal component analysis biplots describing the structure and similarities within the dataset.

in A, whose consumption patterns are similar. Given that the comparison is over 60 weeks at a half-hour granularity, the large extent to which the consumers cluster together in the biplot was unexpected, and indicates that most of the consumers in the dataset have highly similar consumption patterns.

In Fig. 5.10(a), we used the labels in the CER dataset to distinguish the points in the biplot by consumer type. These labels revealed that most residential consumers clustered together in the 2D space, indicating that their consumptions were similar to each other. However,

SMEs varied greatly, which might reflect the unique electricity consumption requirements of their businesses.

In order to capture the relationship among consumers' individual patterns across different weeks, we reshaped A to get another matrix B; it contains $M_B = 48 * 7 = 336$ rows (one for each half-hour of the week) and $N_B = 2,982 * 60 = 178,920$ columns (one for each week of each consumer in the 60-week period). Again, we reduced the dimension of each week from $M_B = 336$ dimensions to 2 dimensions, retaining 68% of the variance in B as shown in Fig. 5.9. The corresponding principal component matrix, P_B, is 2×336, and the PCA score matrix, $Y_B = P_B B$, is $2 \times N_B$.

Note that although A and B contain the same number of elements, their principal component scores are of different dimensions and describe completely different characteristics of the data. The scores in Y_B tell us how similar the 60 weeks of consumption are in the training set across all consumers, and they are plotted in Fig. 5.10(b). We observe a dense clustering of points corresponding to each consumer in the Y_B matrix, which captures how similar the consumption weeks are for each consumer, in comparison to weeks of other consumers. That clustering can easily be seen in Fig. 5.10(b), where we have colored the points corresponding to four very different consumers and their consumption weeks.

A closer look at the weeks for Consumer 1028 in Fig. 5.10(b) revealed a single blue triangle at around $(70, -15)$ in the plot that is significantly distant from the others. It corresponds to Week Index 23 in Fig. 5.1, which is clearly anomalous and probably a vacation week. There are other anomalous points that are distant from the dense cluster. As the anomalies are inherent in the dataset, we assume that they are natural anomalies, and not the consequence of attacks. Attacks, which modify the consumption signals in a manner that changes their pattern, cause a shift in the location of the original point (corresponding to a week) to a completely new one on the biplot, as shown in Fig. 5.11(b).

5.6.3 Clustering Points in the Principal Component Space

A natural density-based clustering of points in the principal component space can be seen in Fig. 5.10. For that reason, we employed the Density-Based Spatial Clustering of Applications

with Noise (DBSCAN) algorithm [102] to determine which points correspond to regular weeks and which points correspond to anomalous weeks. An inherent benefit of DBSCAN is that it works well for irregular geometries of dense clusters, and that it does not assume any underlying probability density of the points. The non-convex boundaries and treatment of dense clusters make DBSCAN better suited to our application than other hierarchical, centroid-based, and distribution-based clustering methods.

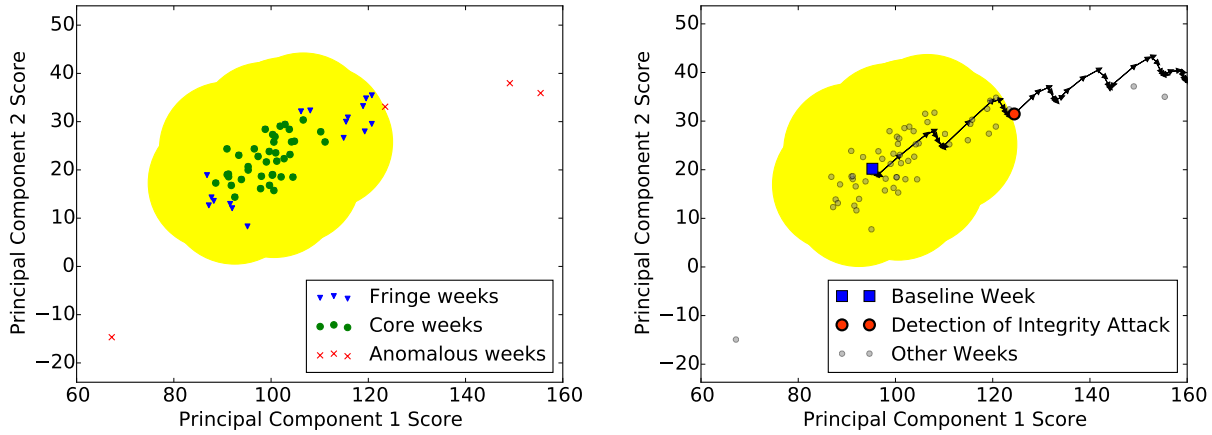
The DBSCAN algorithm has two configurable parameters: *eps* and $MinPts$. These are used to obtain dense neighborhoods. In a two-dimensional Euclidean space, such as our principal component space, the circular region of radius *eps* centered at a point is referred to as the *eps neighborhood* of the point. A *core point* is a point that contains $MinPts$ points within its *eps* neighborhood. All points that lie within the *eps* neighborhood of a core point are considered members of a dense cluster.

In our specific case, we are clustering 60 points that correspond to the weeks of consumption for each consumer in our training set. These points were extracted from Y_B. We define $MinPts$ as the number of points that achieves a simple majority (which in this case is 31). As a result, a single continuous cluster corresponding to normal weeks is produced because any two *eps* neighborhoods containing $MinPts$ points must intersect at at least one point.

Points that lie within the *eps* neighborhood of a core point, but are not core points themselves, are referred to as *fringe points*, as they usually lie at the fringes of the dense neighborhood. The algorithm considers all points that are neither core points nor fringe points to be "noise." This noise is how we define anomalous points in the dataset.

Figure 5.11(a) illustrates the result of the DBSCAN algorithm for Consumer 1028. The green points are core points, and the blue triangles are fringe points. Since we chose *MinPts* to represent a simple majority, the circular *eps* neighborhoods overlap to form a single dense region, indicated by the yellow region in Fig. 5.11(a). All points within this region are considered to be normal. Anomalies, which may be caused by attacks, are points that do not lie in this region. The red crosses correspond to natural anomalies, which the algorithm would flag as false positives, as they lie outside the yellow "safe" region.

Clearly, the detection takes place in the 2D space spanned by P_B's basis vectors, which span the weekly consumptions of all consumers. In order to reverse-engineer the PCA



(a) Decision boundary and types of points in the DBSCAN clustering results

(b) Movement of baseline point with increase in attack duration

Figure 5.11: Principal component analysis biplots for Consumer 1028 capturing (a) the decision boundary for anomalous points and (b) the movement of a baseline week of consumption in the principal component space with increase in duration of an integrity attack.

detector for the purpose of circumventing it, Mallory would need to recreate this 2D space by gaining access to the meters of all consumers. In contrast, she would only need to compromise the smart meter of a single victim in order to reverse-engineer the Average Detector for that victim. *Therefore, the PCA-based detection method is more secure because circumventing it requires full knowledge of all consumers' smart meter readings.*

Although the DBSCAN authors provide recommendations in [102] on how to set *eps*, these methods are not scalable. Specifically, they suggest calculating a list of core distances for each point and observing a knee-point at which a threshold should be set for *eps*. Given that there are $2,982$ sets of points in our dataset (one for each consumer), eyeballing knee-points for each set is not feasible, so we needed to find an alternative. OPTICS, described in [103], can be used to determine cluster memberships for a single set of points that contains multiple clusters. However, it is not suitable for our study, in which we are defining a single cluster per set in $2,982$ sets of points.

We set *eps* based on S_n, a measure proposed by statisticians in [104]. S_n looks at a typical distance between points, which makes it a good estimator of *eps*. In contrast, the median absolute deviation (MAD) and the Mahalanobis distance measure the distance between points and a centroid, which is not how *eps* is defined. And, unlike the standard deviation,

S_n is robust to outliers. In addition, S_n is applicable to asymmetric geometries of points, like those in 5.10(b).

5.7 Detection Using Kullback-Leibler Divergence

The Kullback-Leibler divergence (KLD) is useful for comparing the distribution of a set of measurements against the distribution of a baseline model. This method does not assume any underlying parametric distribution. It is ideal for detecting the integrated ARIMA attack because it can detect the change in consumption pattern distributions caused by the attack. The implementation and evaluation of this method in the context of electricity theft is a main contribution of this chapter.

KLD is calculated as follows. For each consumer, we construct a training matrix X with M rows (one for each week in the M-week training set period). We then use $|B|$ equal bins to calculate a histogram of all values of X, where B denotes the set of bins. We refer to the corresponding probability mass function as the X *distribution*. We iterate through each row i of X (denoted by X_i, where $i \in \{0, 1, ..., M-1\}$), and calculate the probability that values in X_i will lie in each bin of B. Those probabilities give us what we call the X_i *distributions*. It is essential to use the same set of bins, B, determined from the X *distribution* while calculating the X_i distributions. We construct a vector K of KLD measures from a consumer's training set. Each element K_i in K is equal to the KLD between X_i and X:

$$K_i = \sum_{b \in B} P_{X_i}(b) \log_2 \frac{P_{X_i}(b)}{P_X(b)}, \tag{5.12}$$

where $P_{X_i}(b)$ refers to the number of values in X_i that belong to bin b, normalized by the total number of values in X_i. Similarly, $P_X(b)$ is the corresponding relative frequency for values in X that belong to bin b. We refer to the distribution of K as the *KLD distribution*.

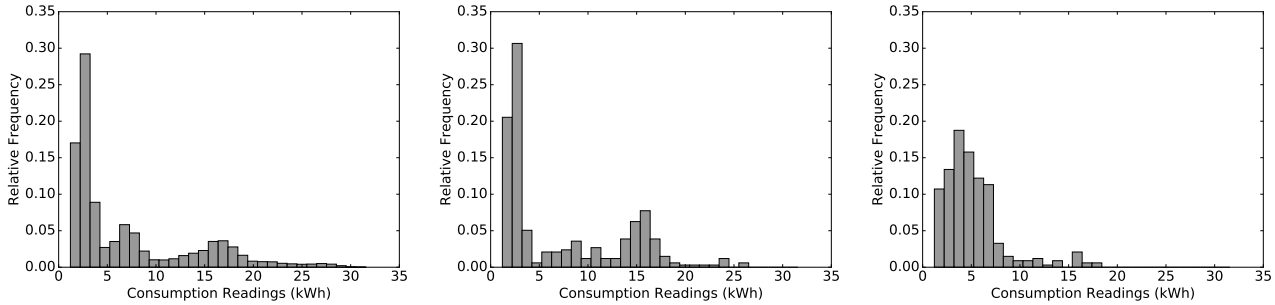
A week of consumption readings is deemed to be anomalous if it deviates too much from the historic distribution. We set thresholds on the KLD distribution at the 90^{th} percentile and 95^{th} percentile for the sake of illustration. For a new week of consumptions, we construct a week vector X_A, and calculate K_A with respect to X using Eq. (5.12). Let the null

hypothesis be defined as the event that a new week of consumption readings, given by X_A, is not anomalous. If $K_A > 90^{\text{th}}$ percentile, we say that the week was anomalous, or the null hypothesis was rejected at an upper-tail significance level of $\alpha = 10\%$. The $\alpha = 10\%$ is a more aggressive detection boundary than the $\alpha = 5\%$ corresponding to the 95^{th} percentile. The reason is that more values are likely to be flagged as anomalous. However, $\alpha = 10\%$ is also subject to a higher false-positive rate, as normal consumptions may also be flagged as anomalous. A successful detector has both a high attack detection rate and a low false-positive rate. The Jensen-Shannon divergence may similarly be used as an alternative to the KL divergence. However, we leave its evaluation for future work.

5.7.1 Illustration of the KLD Detector

In Fig. 5.12(a), we illustrate the baseline X distribution for Consumer 1330 in our dataset. There is one X_i distribution for each week in the training set; we illustrate week 10 (X_{10}) in Fig. 5.12(b). Most training weeks would have similar shapes, while anomalous weeks would have different shapes. For example, the distribution for the integrated ARIMA attack vector is significantly different from the X distribution, as shown in Fig. 5.12(c). The difference between each X_i distribution and the X distribution is given by the KL divergence K_i as defined in Eq. (5.12). K_i is lower for X_{10} than it is for the attack distribution because X_{10} more closely resembles the attack distribution.

The distribution of K_i is illustrated in Fig. 5.13 with the 99^{th} percentile point marked. For this illustration, we used 30 bins. In general, fewer bins produce fewer true and false positives. In the extreme case, a single bin produces 0% true-positive and false-positive rates. Similarly, a very large number of bins could produce higher true-positive and false-positive rates. A more detailed study of the impact of the number of bins on the results is presented in Appendix B.2.



(a) Baseline Training Distribution: X (b) Non-malicious Week Distribution ($K_i = 0.22$) (c) Attack Week Distribution ($K_i = 0.74$)

Figure 5.12: Comparison of the distributions of baseline, non-malicious consumption, and attack readings.

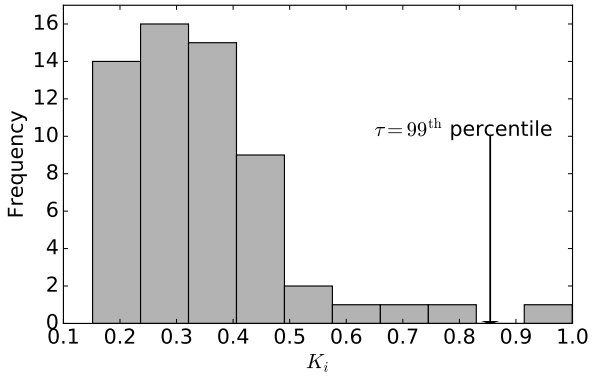


Figure 5.13: Distribution of K_i values for all the weeks in the training set. The 99^{th} percentile is obtained from this distribution.

5.8 Evaluation of the PCA-DBSCAN and KLD Detector on the Min-Average Attack

We compare the performance of our detectors against the min-average detector, which is a heuristic method that was proposed in related work [86]. We show that our detectors dramatically mitigate the min-average attack. We also evaluated the detectors against the integrated ARIMA attack, which we proposed, but include that in Appendix B.

For each of the 500 consumers in the CER dataset, we estimated how much Mallory would be able to gain in one week through the use of the min-average attack because of detector false negatives. In that sense, we allowed Mallory to take the roles of any of the 500 consumers, taken one at a time. We separately evaluated the PCA-DBSCAN and the

KLD detector on all 500 consumers, and set the most aggressive detection thresholds for each consumer such that the false positive rate was at most 7% on the 60 weeks of training data. We implemented the min-average attack and marked positive gains when the min-average attack was not detected by either of the two detectors. Those gains were computed by injecting the attack into a test set of 14 weeks of data, and computing the difference between Mallory's true bill (the bill she would have received if she had not committed fraud) and fraud bill (the bill she would receive having committed fraud). We averaged those gains over the 14 weeks to obtain the gains for one week. Figure 5.14 summarizes the results for both detectors. It can be seen that, for most consumers, both detectors are able to detect the attack. Therefore the gains are zero for most consumers. The gains were large for a few consumers for which the detectors could not perform well because of consumption behavior unpredictability.

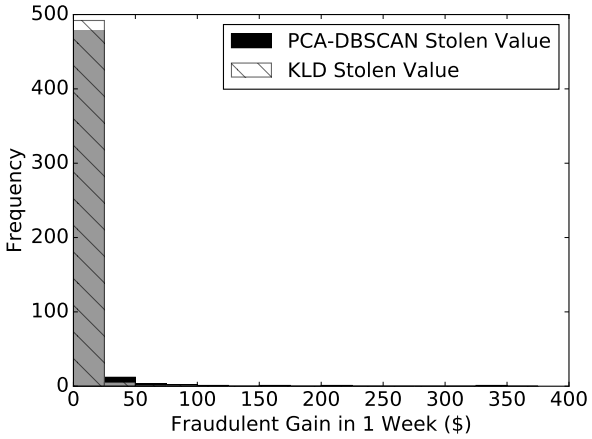


Figure 5.14: Distribution of the amount of electricity that can be stolen in one week through the use of the min-average attack against the PCA-DBSCAN detector ($3.8 on average) and the KLD detector ($1 on average).

From Fig. 5.14, it can be seen that the KLD detector mitigates the min-average attack. In the absence of that detector, the gains would have been $26.4 per week (as described in Section 5.4.2); with the detector in place, the gains were only $1 per week. In summary, on average, the KLD detector mitigates the min-average attack by 96%. Similarly, the PCA-DBSCAN detector mitigates the gains by 86% from $26.4 per week to $3.8 per week.

5.9 Related Work

In [98], the authors describe how they simulated consumption patterns of loads in households and detected changes in these patterns to identify electricity theft achieved by tapping power lines. They reduce false-positive rates by fusing alerts reported by multiple sensors. Their approach is similar to ours in that they try to identify ways in which attacks can take place, and employ learning algorithms to detect attacks. The authors of [78], [98], and [105] all independently claim to have built comprehensive attack trees that span all possible electricity theft attacks. However, their attack trees all depend on existing technologies. In contrast, our work analyzes the fundamental necessary conditions for the execution of a successful electricity theft attack, which are equally applicable to future, unknown technological approaches for such attacks.

In this chapter, we contribute to the state of the art and the state of the practice by evaluating cyber-based electricity theft attacks on the smart grid in the presence of security measures that are currently employed by industry. We are the first to employ and study the Kullback-Leibler (KL) divergence in the context of detecting electricity theft, but this method has been used in other anomaly detection applications in [106] and [107].

5.10 Conclusion

In this chapter, we characterized anomaly detection methods for time-series data based on whether they detect anomalies in either individual readings or sets of readings. We discussed the min-average detector from related and work and derived the optimal attack against that detector. We proposed four fraud detectors, three of which detect anomalous sets of readings. We proposed the ARIMA detector and the integrated ARIMA detector to improve on the ARMA detector from related work. We proposed the PCA-DBSCAN detector and explained its mathematical underpinnings. We proposed the KLD detector and explained how it could be applied in practice.

We evaluated the PCA-DBSCAN detector and the KLD detector against the min-average detector (proposed in related work), and found that the attacker's monetary gains via fraud

were reduced by 96% through the use of the KLD detector and 86% through the use of the PCA-DBSCAN detector. Therefore, we demonstrated the claim in our thesis statement that empirical models can improve fraud detection in the context of consumption fraud. The work in this chapter was peer-reviewed, and published in [33], [47], [89], and [95].

In the next chapter, we will evaluate the detection methods that were proposed in this chapter against attacks that are optimal against them.

CHAPTER 6

OPTIMAL ATTACK VECTORS THAT ENABLE CONSUMPTION AND GENERATION FRAUD

In Chapter 5, we presented the min-average attack, which is the optimal attack against the min-average detector (worst-case scenario for the detector). By *optimal*, we mean that Mallory's gains are maximized while remaining within the detection threshold and circumventing the detector. We evaluated the PCA-DBSCAN and KLD detectors against the min-average attack. Although the min-average attack is optimal against the min-average detector, it is not optimal against the PCA-DBSCAN and the KLD detectors. In this chapter, we will take a different evaluation approach; we will evaluate each detector based on the optimal attacks against that detector. In doing so, we compare detectors on how well they would perform if the attacker knew how they were designed.

Similar to consumption fraud discussed in Chapter 5, smart grid entities with generation capabilities can compromise their meters to over-report their generation to make monetary gains. For distributed energy resources (DERs) like solar and wind power, in particular, such compromise may be advantageous. That is because their capital costs of installation are high. Although those costs have been decreasing in recent years [16], it can take a long time for DER owners to recover them just through income from generation. In this chapter, we show that there is a compelling motivation for DER generation fraud because it can reduce the time it takes to recover the capital costs of DER installation by over 80%.

The adoption of renewable energy generators has been increasing rapidly in recent years. From 2010 to 2015, photovoltaic adoption in the U.S. grew by 46%, 43%, and 101% for residential, commercial, and utility-scale installations, respectively [108]. Globally, solar capacity increased by 28% and wind capacity increased by 17% from 2014 to 2015 [16]. Most capacity additions in the U.S. came from wind power (41%) in that period [17], and by the end of 2016, wind surpassed hydro as the largest source of renewable energy in

the U.S. [109]. The worrying trend is that demand for electricity has not grown with the generation, and as recently as April 2017, it was reported that wholesale electricity prices had dropped so low that at times they were even negative [110]. Since the operating costs of distributed energy resources (DERs) such as solar and wind (about \$15/MWh [111]) often exceed the amount that DER owners are paid (ranging from \$0 to \$45/MWh [110]), there is a real motivation for DERs to compromise their reported generation in order to make fraudulent monetary gains and ensure that they profit.

In this chapter, we present the first analysis of how much attackers, who own or operate DERs, would stand to gain by fraudulently reporting that they generate more than they actually do. The detection of DER fraud, in the case of solar and wind, is different from that of consumer fraud in that DERs generate electricity based on weather conditions. The dependence of DERs on weather creates correlations between nearby DERs that can be leveraged for detection.

In considering DERs, we restrict our attention to solar and wind, which are the most prevalent sources. We refer to customers who seek to make monetary gains through consumption or generation fraud as *attackers*. The attackers may be individuals or groups of individuals who own or operate electricity consumption facilities and/or DERs. Consumers who also produce electricity are referred to as *prosumers*.

6.1 Summary of Contributions

In this chapter, we derive the optimal attacks against the KLD and PCA-DBSCAN detectors in the context of consumption fraud in Section 6.2. We present the first study of detectors for mitigating DER fraud. The detectors analyze smart meter readings collected at a central location, which is presumably at the utilities' data centers. The detectors then construct an unsupervised model of normal consumption and generation behaviors. The models of normal consumption/generation are unsupervised because there are infinitely many ways in which attackers could compromise their meter readings in order to make monetary gains, so labeling of attacks for supervised modeling cannot be done in a comprehensive manner. In this chapter we present modeling approaches that are unique to DERs, and leverage

correlations between DERs and relevant weather data. The anomaly detection methods leverage those models to determine whether or not meter fraud has taken place.

We describe the datasets used in both our illustrations of generation fraud and our evaluations of fraud detectors in Section 6.3. We propose fraud detectors and derive optimal attacks against those detectors in Section 6.4. We present a unique study of the costs of solar and wind installations in Section 6.5, and show that those optimal attack vectors can dramatically reduce the time it would take for a generator owner to recover those costs. We conclude the section in Section 6.6.

6.2 Optimal Attack Vectors against Detectors of Consumption Fraud

An optimal attack would maximize the amount of electricity stolen while going undetected. In this section, we derive the optimal attacks against the KLD detector and the PCA-DBSCAN detector in the context of consumption fraud.

6.2.1 Optimal Attack Against the KLD Detector

In order to construct the optimal attack against the KLD detector, the attacker (denoted by A) would need to have access to three pieces of information: (1) the set of bins being used, B; (2) the distribution of the training data, D'_{Tr}, over those bins; and (3) the percentile threshold calculated on the training data, τ. The optimal attack vector for a week of readings, D'^{*}_{A}, maximizes the attacker's profit as follows.

$$D'^{*}_{A} = \arg\max_{D'_A} \sum_{t=1}^{T} \lambda(t)\Delta t[D_A(t) - D'_A(t)] \tag{6.1}$$

$$\text{subject to } \mathrm{KLD}(D'_A, D'_{Tr}) \leq \tau, \tag{6.2}$$

where $\lambda(t)$ is the electricity price at time t and Δt is the time period between the collections of readings. The objective function is the profit given in dollars ($); Δt is in hours (h); $D_A(t)$ is in kilowatts (kW); and $\lambda(t)$ is in $/kWh. The space of readings is very large (but

countably finite because readings are rounded off to the nearest watt and bounded below by zero). Therefore the search space is exponentially large in T. However, we show that the problem can be reformulated as a mixed-integer convex optimization problem and efficiently solved using free solvers. The trick is as follows. Instead of solving for D'^{*}_{A}, we solve for the optimal distribution of D'^{*}_{A} such that the KLD from the training data remains within the threshold τ. Once we have the optimal distribution, we can generate D'^{*}_{A} as per that optimal distribution.

We can restate the objective function in Eq. (6.2) as follows by removing all the constants.

$$D'^{*}_{A} = \arg\max_{D'_A} \sum_{t=1}^{T} \lambda(t)[D_A(t) - D'_A(t)] \tag{6.3}$$

$$= \arg\min_{D'_A} \sum_{t=1}^{T} \lambda(t) D'_A(t) \tag{6.4}$$

$$= \arg\min_{D'_A} \frac{1}{T} \sum_{t=1}^{T} D'_A(t), \tag{6.5}$$

where the last equality assumes that $\lambda(t)$ is constant for all t (we will later relax that assumption). Therefore, by minimizing the mean of the reported readings, we can maximize the profits for the attacker. Let the probability distribution of $D'_{Tr}(t)$ be discretized into $|B|$ bins with bin centers $X_{Tr}(b)$ for $b \in B$. Let $P_A(b)$ denote the probability $Prob(D'_A(t) \in b)$ for $b \in B$. Then the mean given in Eq. (6.5) can be expressed in terms of the expectation of the probability distribution as follows: $\frac{1}{T}\sum_{t=1}^{T} D'_A(t) = \sum_{b\in B} X_{Tr}(b) P_A(b)$. With that formulation, we can use $P_A(b)$ as the optimization variable, and that is convenient because the KLD constraint can be expressed directly in terms of $P_A(b)$ and the corresponding training probabilities $P_{Tr}(b)$. The equivalent convex optimization formulation of the optimal attack

against the KLD detector is given as follows.

$$P_A^* = \arg\min_{P_A} \sum_{b \in B} X_{Tr}(b) P_A(b) \tag{6.6}$$

$$\text{subject to } \sum_{b \in B} P_A(b) \log \frac{P_A(b)}{P_{Tr}(b)} \leq \tau, \tag{6.7}$$

$$\sum_{b \in B} X_{Tr}(b) P_A(b) \geq \text{min\_avg}, \tag{6.8}$$

$$P_A(b) \geq 0, \tag{6.9}$$

$$\text{and } \sum_{b \in B} P_A(b) = 1. \tag{6.10}$$

The first constraint ensures that the KLD value is less than the threshold τ, the second constraint includes the min-average detector, and the last two constraints ensure that the probability values are valid. X_{Tr}, P_{Tr}, and τ are constants. The objective function is a linear sum, and the probability validity constraints are also linear. The KLD constraint is convex, making the whole problem a convex optimization problem that solves for the $P_A(b)$ values from which the attack vector can be generated. Since the probabilities need to be converted to integer frequencies, which can be used to generate the attack vector, the problem becomes mixed-integer non-linear program (MINLP). In general MINLP is NP-Hard, but when the constraints are convex, an exact solution can be obtained in polynomial time. We obtained the solution using the Bonmin solver [112]; alternative solvers are described in [113].

An example of an optimal attack for one particular consumer in the CER dataset is illustrated in Fig. 6.1 with $|B| = 10$ and τ set at the 90^{th} percentile of the KLD values in the training set. Notice that the attacker is trying to under-report consumption, as evidenced by the fact that the bin corresponding to the lowest consumption readings has a probability associated with it in the attack distribution that is higher than in the training distribution. Also, the bins corresponding to larger consumption readings have lower probabilities associated with them. The smallest consumption value in each bin is, by design, the value at the left edge of the bin; $D_A'^*$ can be generated by making copies of the smallest value in each bin in a manner that adheres to the optimal distribution.

The above derivation assumes that the electricity price $\lambda(t)$ is a constant, as in flat-rate

pricing. For time-of-use pricing, the optimal attack vector is more tedious derive, but the underlying principles are the same as they are in the case of flat-rate pricing. Assume that there are two prices (the idea can be extended for more than two prices), peak (denoted by λ_p), and off-peak (denoted by λ_{op}). Furthermore, in a period of one week, let us assume that there are T_p peak periods and T_{op} off-peak periods. In this case, we need to obtain $D'^*_A = [D'^*_{A\_p}, D'^*_{A\_op}]$ denoting the reported readings for the peak and off-peak periods. Note that the ordering of readings D'^*_A within the peak and off-peak periods is irrelevant because the price is constant during those periods. The optimization objective in Eq. (6.5) is modified for peak and off-peak pricing as follows.

$$[D'^*_{A\_p}, D'^*_{A\_op}] = \arg \min_{[D'_{A\_p}, D'_{A\_op}]} [\lambda_p \sum_{t=1}^{T_p} D'_{A\_p} + \lambda_{op} \sum_{t=1}^{T_{op}} D'_{A\_op}] \tag{6.11}$$

$$= \arg \min_{[D'_{A\_p}, D'_{A\_op}]} [\lambda_p T_p \mathrm{E}[D'_{A\_p}] + \lambda_{op} T_{op} \mathrm{E}[D'_{A\_op}]]. \tag{6.12}$$

The above objective can be translated into a probabilistic expectation so that the objective function and the constraints use the same optimization variable, similar to Eq. (6.10). That is done by obtaining two histograms $P^*_{A\_p}$ and $P^*_{A\_op}$ describing the optimal distribution of readings within the peak and the off-peak periods, respectively. Note that the bins for both histograms are given by the same set B. The corresponding optimization problem is

formulated as follows.

$$[P^*_{A\_p}, P^*_{A\_op}] = \arg \min_{[P_{A\_p}, P_{A\_op}]} [\lambda_p T_p \sum_{b=1}^{B} X_{Tr}(b) P_{A\_p}(b) + \lambda_{op} T_{op} \sum_{b=1}^{B} X_{Tr}(b) P_{A\_op}(b)] \quad (6.13)$$

subject to (6.14)

$$P_{A\_p} \geq 0 \; ; \; \sum_{b=1}^{B} P_{A\_p} = 1 \quad (6.15)$$

$$P_{A\_op} \geq 0 \; ; \; \sum_{b=1}^{B} P_{A\_op} = 1 \quad (6.16)$$

$$\sum_{b=1}^{B} X_{Tr}(b) P_A(b) \geq \text{min\_avg} \quad (6.17)$$

$$\sum_{b \in B} P_A(b) \log \frac{P_A(b)}{P_{Tr}(b)} \leq \tau, \quad (6.18)$$

where

$$P_A(b) = \frac{T_p P_{A\_p}(b) + T_{op} P_{A\_op}(b)}{T_p + T_{op}}. \quad (6.19)$$

The above objective function encodes the objective function in Eq. (6.12). The first two constraints ensure that $P^*_{A\_p}$ and $P^*_{A\_op}$ are valid probability distributions. The third constraint is the min-average detector constraint and the final constraint is the KLD detector constraint.

As a consequence of solving the above optimization problem, obtaining the optimal distributions, and generating the optimal attack vector D'^*_A from those distributions, the attack injects smaller values when the electricity price is low, and larger values when the electricity price is high. That is illustrated in Fig. 6.2, in which TOU pricing corresponding to the CER dataset was applied, and larger values in the optimal attack vector were injected during off-peak periods, when the price was low. As a result, the attacker was charged less for larger consumption values.

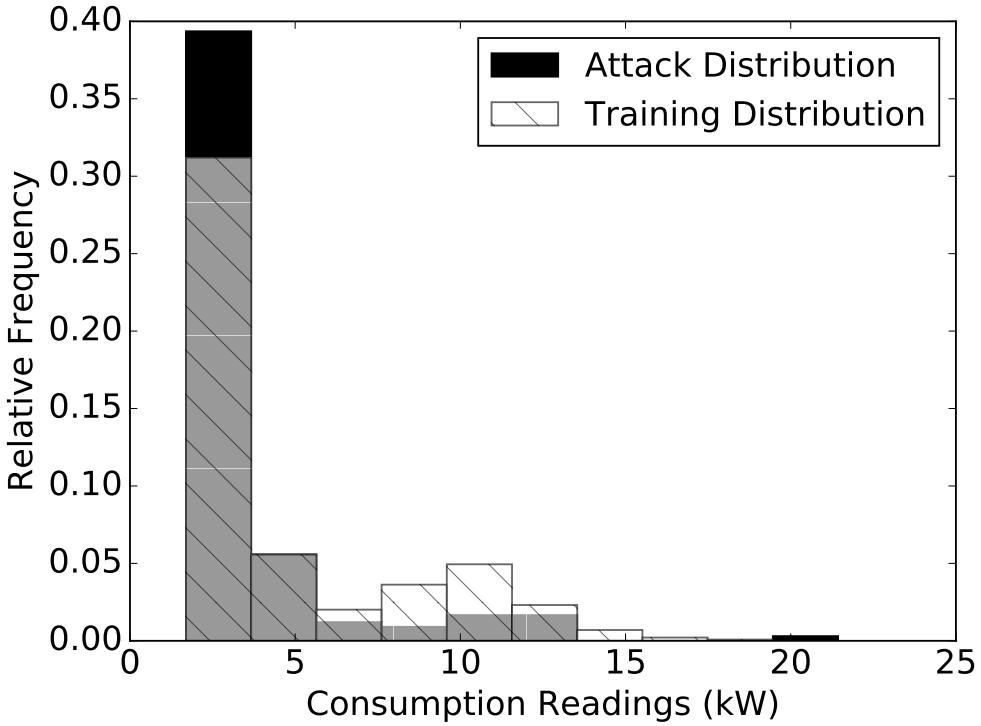


Figure 6.1: Distribution of the optimal attack against the KLD detector in comparison to the training distribution.

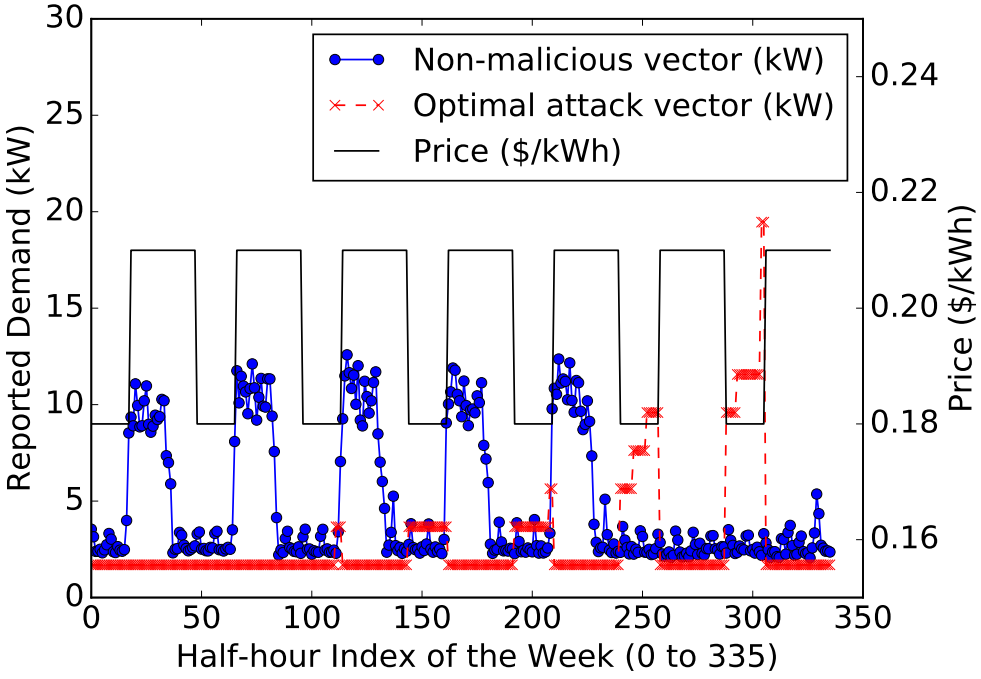


Figure 6.2: Illustration of optimal attack against KLD with TOU pricing.

6.2.2 Optimal Attack Against the PCA-DBSCAN Detector

In principle, the PCA-DBSCAN detector is very similar to the KLD detector. M weeks in the training set, each containing N consumption readings, are transformed into an R-dimensional space, where $R << N$. In the CER dataset, $N = 336$. Let V be an $N \times R$ matrix that performs the PCA transformation $Y = XV$ of the training set X. Note that X is $MC \times N$ and contains the training data for all C consumers, and V is calculated on that full dataset. For the sake of obtaining the detection boundary, the training data for each consumer, denoted by D'_{Tr}, are independently projected into the R-dimensional space by using V. V, however, is calculated on X, which combines the data for all consumers.

As per the approach in [89], the DBSCAN algorithm finds a subset of the M training weeks that are not anomalous in the vector space of Y. First, it determines *core weeks*, which are points in Y for each consumer that contain $M/2$ neighbors in Y that are within an ϵ radius (as measured by the Euclidean $L2$ norm). Any points in Y that lie outside of the ϵ radius of all core weeks are deemed anomalous, and indicative of an attack. An example for a consumer in the CER dataset is illustrated in Fig. 6.3, in which the core weeks are projected from a 336-dimensional space onto a 2-dimensional space. The *safe* region is shaded in yellow and contains overlapping circles centered at core weeks with radius ϵ. All points outside that region are marked as attacks. For example, the *zero attack* vector, which sets all consumption readings to zero, is marked as an attack because its projection lies far outside the safe region.

Let η denote the set of core weeks, a subset of the M training weeks for each consumer, which is determined by DBSCAN. In the example illustrated in Fig. 6.3, there are $M = 60$ weeks in the training set, of which $|\eta| = 52$ are core weeks. The remaining 8 weeks are not as typical as 52 core weeks, and some of them may be anomalous. Therefore, they are not used to model normal consumption behavior.

The optimal attack vector contains $M = 336$ readings in one week and must project into the safe region. The objective is taken from Eq. (6.5), and the problem is formulated as

follows.

$$D'^{*}_{A} = \arg\min_{D'_A} \sum_{t=1}^{N} \lambda(t) D'_A(t) \tag{6.20}$$

$$\text{subject to } D'_A \geq 0, \tag{6.21}$$

$$D'_A \leq max(D'_{Tr}), \tag{6.22}$$

$$\frac{1}{T} \sum_{t=1}^{T} D'_A(t) \geq \text{min\_avg} \tag{6.23}$$

$$||D'_A V - D'_n V|| \leq \epsilon, \forall n \in \eta, \tag{6.24}$$

where $\lambda(t)$ would be known beforehand in flat-pricing or time-of-use schemes. $D'_A \leq max(D'_{Tr})$ is an upper-bound constraint imposed based on the maximum of the historic data in the training set. The distance between the projected attack, $D'_A V$, and the projected core week, $D'_n V$, is measured by the $L2$ norm, and that constraint is repeated $|\eta|$ times over all the different core weeks. Since the $L2$ norm is a convex function and the objective function is linear, the optimization problem is convex and can be solved efficiently using solvers like SCS [114], which comes packaged with CVXPY [115] for Python.

Note that the optimization problem is solved *separately* for each $n \in \eta$. In other words, $|\eta|$ different minimization problems are solved, and the solution corresponding to the minimum of all minima is the final solution. That is based on the result from real analysis that the infimum of a union of sets is the infimum of the set of infima of the individual sets. In our case, the feasible set is the union of hyperspheres (circles in two dimensions, as illustrated in Fig. 6.3) centered at $D'_n V$ with radius ϵ. If the optimization problem were solved by taking all the constraints $n \in \eta$ together, then the feasible set would have been the intersection of those hyperspheres; that is, by definition, not the same as our detection boundary, which is given as the boundary of the union of the hyperspheres.

An example of the optimal attack in comparison to core weeks is illustrated in the lower-dimensional space in Fig. 6.3 and in the higher-dimensional space in Fig. 6.4. In Fig. 6.3, notice that the optimal attack vector is projected onto the boundary of the feasible set. That point represents maximum gains for the attacker while remaining within the detection

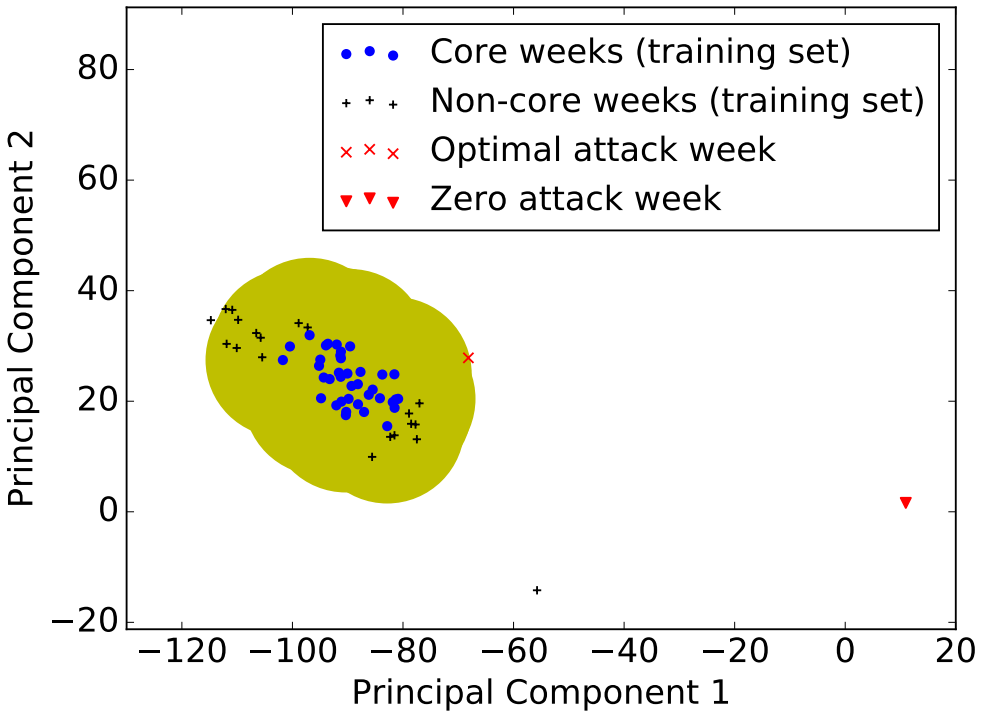


Figure 6.3: Core weeks and attacks projected, using PCA, onto a two-dimensional space. Points outside the yellow region, which was formed by overlapping circles centered on core weeks, are marked as attacks. The optimal attack circumvents detection and lies within the detection boundary.

boundary, and going undetected. In Fig. 6.4, notice that the optimal attack has mostly zeroed values, but the few nonzero values are large and ensure that the projection lies in the safe region. In addition, the nonzero values in the optimal attack coincide with peaks in the consumption readings of the core weeks. That ensures that the optimal attack vector preserves the trend in the training data. In preserving that trend, the optimal attack vector does not adversely affect the low-rank approximation and lies far from the core weeks in the low-dimensional space.

Note that the optimal attack vector against the PCA-DBSCAN detector incorporates the timing of reported meter readings with the TOU electricity prices. It is different from the optimal attack vector against the KLD detector, which is agnostic to the time-ordering and allows the freedom to inject larger consumption values when the price is low. Also notice that both optimal attack vectors require consumption to be over-reported at certain times, in order to avoid detection.

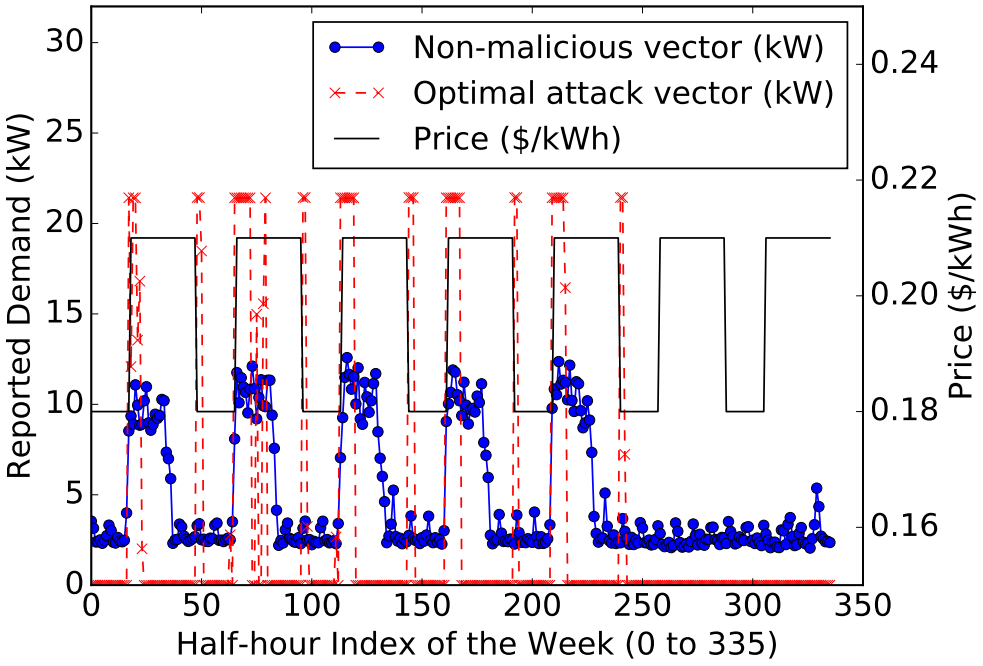


Figure 6.4: Illustration of optimal attack against PCA in the original dimension.

6.2.3 Comparison between KLD Detector and PCA-DBSCAN Detector

We had compared the KLD detector and the PCA-DBSCAN detector in terms of how well they detected the the min-average attack and the integrated ARIMA attack in Chapter 5. In this section, we compare them based on their worst-case scenarios. The comparison was made as follows. For each detector, we allowed Mallory to take the role of all 500 consumers in the CER dataset, one at a time. In each of those roles, Mallory was able to circumvent the detector designed specifically for the consumer. For a fair comparison, the detector threshold was set such that the false-positive tolerance was 7%. Therefore, the detector thresholds for both detectors were set to be as tight as possible while ensuring that no more than 7% of weeks in the training set triggered false positives. For each consumer role played by Mallory, the fraudulent gains were obtained by calculating the difference between the optimal attack D'^{*}_{A} generated (for the two detectors discussed earlier in this section) and the true consumption in a hold-out set of 14 weeks of data. The gains were averaged over the 14 weeks to obtain the average gains per week for each consumer.

The gains in one week for all 500 consumers playing the role of Mallory are plotted as

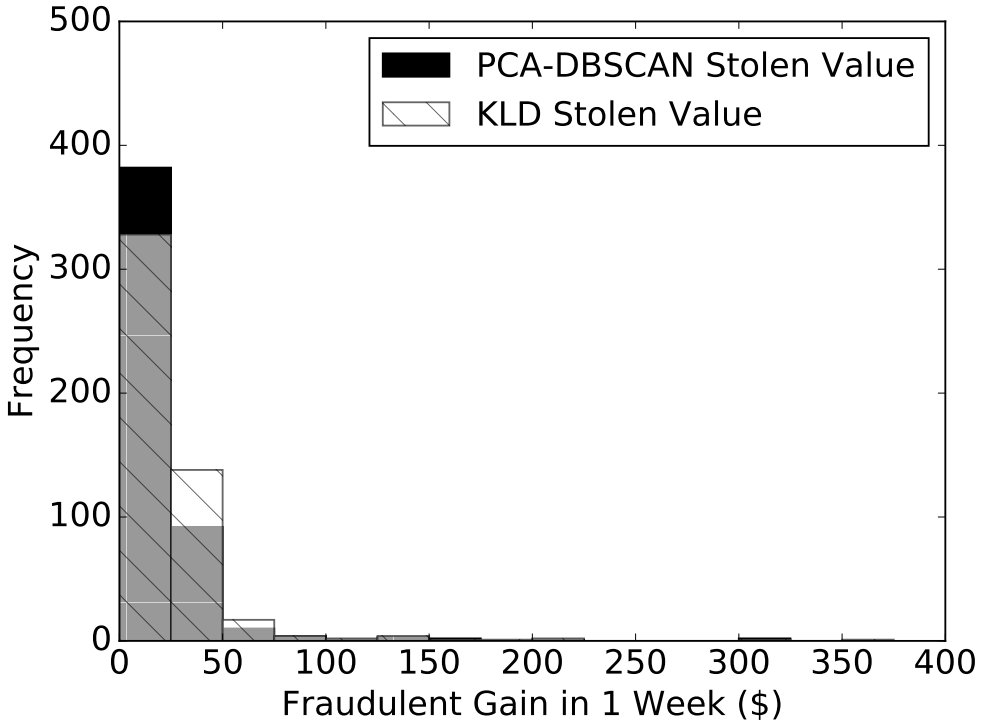


Figure 6.5: Distribution of the amount of electricity that can be stolen in one week through the use of optimal attacks against the PCA-DBSCAN detector ($21.3 on average) and the KLD detector ($24.9 on average).

a histogram in Fig. 6.5. It is clear from the histogram the the PCA-DBSCAN detector outperforms the KLD detector on average.

Note that the average gains in one week from the PCA-DBSCAN detector ($21.3) and the KLD detector ($24.9) are not significantly lower than the gains from the optimal attack against the min-average detector ($26.4); recall that the optimal attack against the min-average detector was obtained in Section 5.4.2. That raises the question of the benefit of the two proposed detectors. The benefit lies in the fact that the optimal attack against the proposed detectors is much harder for Mallory to compute than is the optimal attack against the min-average detector. The reason is that the optimal attack against the min-average detector requires the knowledge of only Mallory's own readings to be able to compute the detection threshold, which is the smallest of the averages of past weeks of readings. The optimal attack against the PCA-DBSCAN detector requires knowledge such as the PCA projection matrix V and the radius ϵ, which can be obtained only with utility-insider

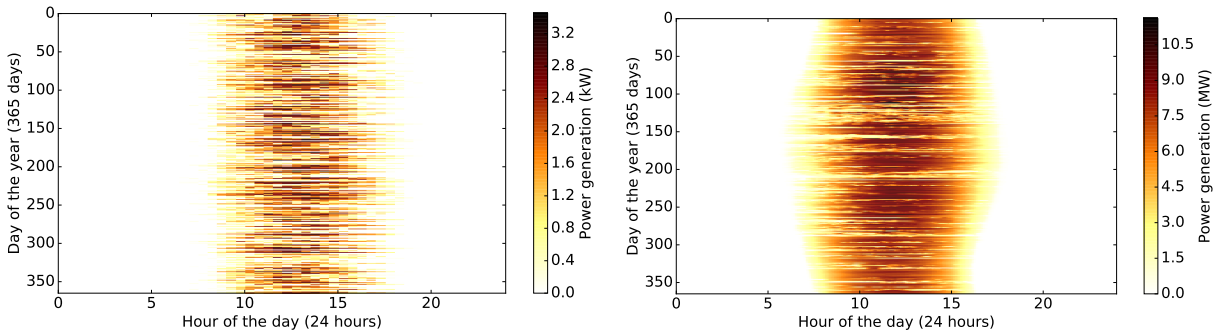
knowledge (the set of core points η can be computed from Mallory's own readings). Similarly, the set of bins B and the KLD threshold τ can be obtained only from a utility-insider to circumvent the KLD detector. Therefore, obtaining the worst-case attacks against the PCA-DBSCAN and KLD detectors would be difficult to achieve for most malicious consumers in practice.

6.3 Description of Datasets used in Designing Detectors of DER Fraud

We use three datasets to evaluate our approaches to mitigating generation fraud. The fact that the datasets are all freely available makes it possible for other researchers to replicate and extend our results. We assume that the datasets have not been compromised by an attacker, and use the data to model normal behavior, from which attack behavior can be distinguished. Note that the datasets may have anomalous behavior that can lead to false positives. The data come from consumers and generators in Australia, France, Ireland, and the U.S.

6.3.1 Ausgrid Solar Dataset

This is an openly available dataset of electricity consumption and generation measurements taken from a real deployment of 300 customers in the Sydney area [116] who have rooftop solar panels on their homes. Readings were taken at a half-hour time granularity for one year. In *net metering*, the primary purpose of the solar panel is to meet the customer's own consumption needs. If consumption exceeds the solar generation, the deficit power is supplied by the utility grid at the prevailing retail price. If generation exceeds consumption, the excess generation is sold back to the utility because the customers do not have storage on their premises. The net generation is illustrated in Fig. 6.6(a).



(a) Ausgrid: Net solar generation

(b) NREL: Gross solar generation

Figure 6.6: Solar generation datasets: Heatmap illustration of daily repeating patterns for one photovoltaic in the Ausgrid dataset (rated at 9 kW) and one photovoltaic in the NREL dataset (rated at 13 MW).

6.3.2 NREL Solar Dataset

This dataset was created by the National Renewable Energy Laboratory (NREL) to be representative of solar output characteristics across the U.S. [117]. We examined the metered generation of 238 distributed photovoltaics in California from the dataset. Those photovoltaics had ratings ranging from 4 MW to 121 MW, and the data for one of the photovoltaics is plotted in Fig. 6.6(b). The data were produced at a 5-minute granularity for a period of one year.

6.3.3 Engie Wind Dataset

This is an openly available dataset of wind power generation from four 2 MW turbines in Meuse, France. The data were provided by Engie, a French utility company. Readings were taken at a 10-minute granularity, and we extracted a period of one year in which all four turbines were continuously operational. One sample turbine group rated at 16 MW is illustrated in Fig. 6.7.

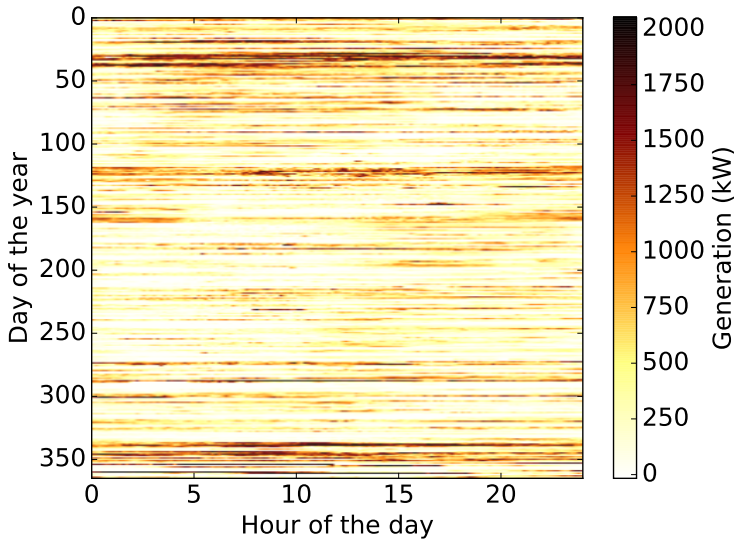


Figure 6.7: Engie wind dataset: Sample utility-scale turbine rated at 2 MW.

6.4 Optimal Attack Vectors against Detectors of DER Fraud

As described in the attack model in Section 4.6, an attacker may commit extensive fraud by over-reporting generation by an arbitrarily large value. In this section, we present detectors that can mitigate such fraud. For each detector, we identify the worst-case attack (optimal for the attacker), which maximizes the electricity stolen while avoiding detection. We evaluate each detector based on how much an attacker can gain by using the optimal attack that circumvents that detector.

6.4.1 Rating Attack

A rating-based detector to limit the over-reporting of attacker generation would ensure that the reported generation does not exceed the DER's rating. The optimal attack for this detector sets the generation readings at the rating threshold. In doing so, the attack vector does not exceed the threshold. Simultaneously, the attackers maximize how much they can steal by over-reporting their generation. This is the optimal attack for a rating-based threshold, and we refer to it in this chapter as the *rating attack*.

In the case of solar, the generation is zero before sunrise and after sunset. Therefore, in designing the rating-based detector for solar generation, we ensure that the detection

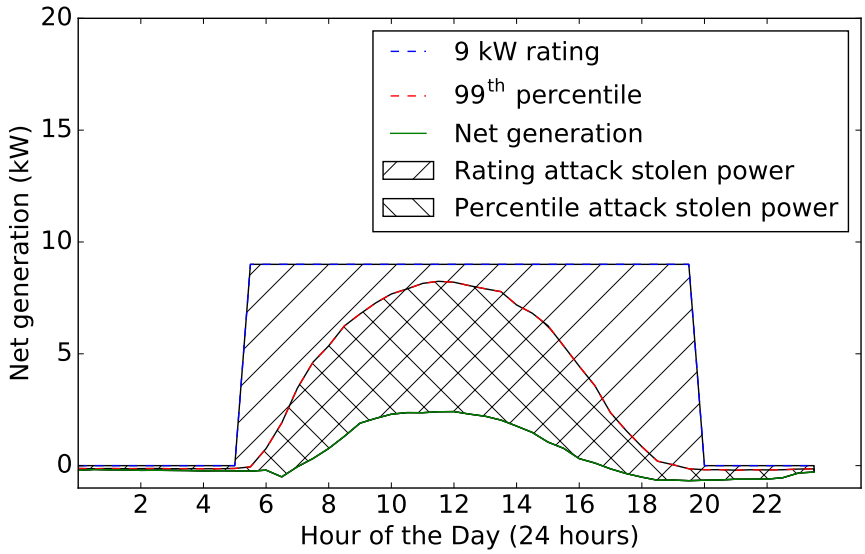


Figure 6.8: Rating and percentile attacks illustrated for one customer in the Ausgrid solar dataset. The shaded regions represent the stolen electricity.

threshold is set such that the generation can never exceed zero during that period. From sunrise to sunset, the upper threshold is the solar panel's rating. The amount of electricity stolen is the difference between the rating and the actual generation, as illustrated in Fig. 6.8.

In a net metering system, if the DER owners were to claim that the net generation was equal to the rating of the panel, they would in effect be claiming that their consumption was zero. In doing so, they would not only over-report their generation, but also under-report their consumption, which is theft.

Unlike solar generation, wind generation of a turbine can reach its rated capacity at any time of the day or night. Therefore, the detection threshold would be set at the rated capacity of the turbine throughout the day.

6.4.2 Percentile Attack

The rating attack, particularly for solar power, is naive in that it does not capture diurnal variations of generation. In solar, for example, the output steadily increases until midday and then steadily decreases in the evening, according to the solar irradiance. In order to determine whether each solar output reading at a particular time is anomalous, one approach

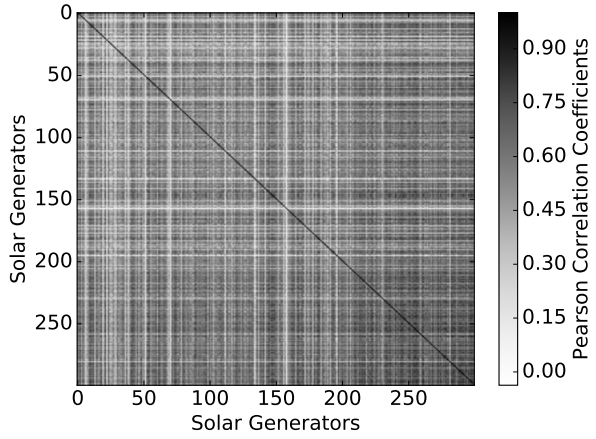
may leverage the diurnal patterns, and compare that reading with readings taken at the same time on previous days. Our percentile threshold accomplishes that by setting a threshold at the 99^{th} percentile of data points seen at the same time on previous days. For example, to determine whether a reading at 10:00 A.M. on a given day is anomalous, we check whether that reading is greater than the 99^{th} percentile of generation values taken at 10:00 A.M. on previous days. Our choice of percentile point is based on the desire to achieve an acceptable trade-off between true positives and false positives.

We refer to the optimal attack against the detector that uses the percentile-based threshold as the *percentile attack*. Launching the percentile attack requires that the attacker be aware that the utility may be using that detector as a defense. It also requires the attackers to build a model of their own generation, in order to circumvent the detector by setting their generation at the 99^{th} percentile threshold. In doing so, they would not exceed the threshold, and would maximize how much they can steal by over-reporting their generation while going undetected. The amount of electricity stolen is the difference between the percentile threshold and the actual generation, as illustrated in Fig. 6.8. Since that amount is always less than that of the rating attack, the percentile-based threshold mitigates the extent of possible fraud, relative to the rating-based threshold.

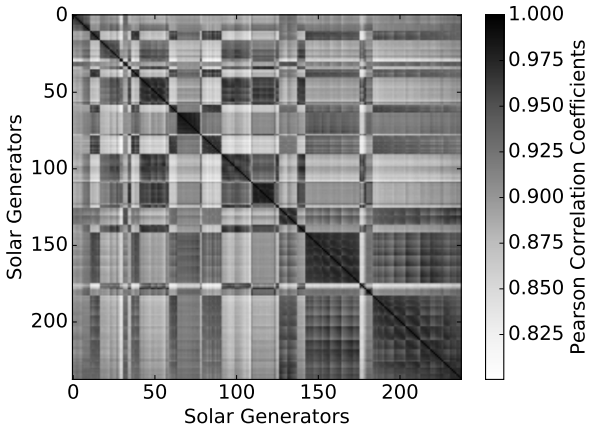
6.4.3 Correlation Attack

One way to detect attacks in the context of DERs, like solar and wind, is to leverage their dependence on the availability of sunlight and wind, respectively. Therefore, the generations of different DERs would be expected to be correlated. We verified that with all three DER datasets; the results are illustrated in Fig. 6.9. The heatmaps in the figure were obtained by computing the pairwise Pearson correlation coefficients between the DERs in the datasets.

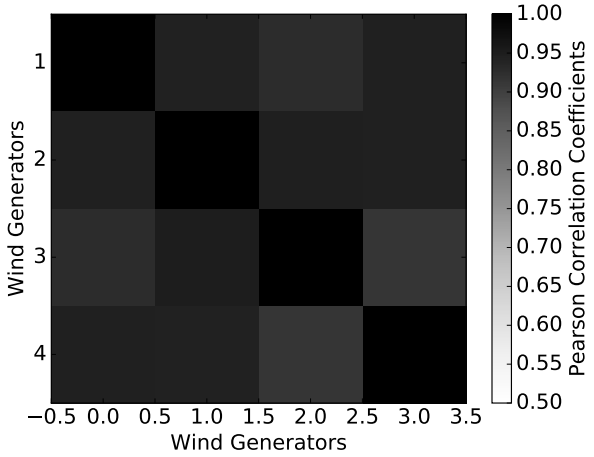
For the NREL solar dataset plot, shown in Fig. 6.9(b), we had ordered the consumers in increasing order of photovoltaic ratings. We were surprised to find that the cross-correlations were maximum between adjacent generators. That means that generators that had similar ratings were maximally correlated, which is surprising given that the data were standardized and given that the generators were located in different parts of California. This observation



(a) Ausgrid residential-scale solar



(b) NREL utility-scale solar



(c) Engie wind turbines

Figure 6.9: Cross-correlations. These heatmaps plot the Pearson correlation coefficient between all pairs of DERs in the dataset. Note that the minimum values on the scales are not zero, so the cross-correlations are all high.

is useful for defenders (utilities) because the *correlation detector* works by validating the readings of each DER using the readings from the most highly correlated neighboring DER. Finding the most highly correlated DER can be computationally expensive when there are many DERs to choose from, but utilities could reduce the search space by limiting their search to the DERs with similar ratings.

Let A denote the attacker and C denote a neighboring DER used for detection. Per the detector's design, C's correlation with A must be greater than any other DER's correlation

with A. Correlation implies a linear relationship, which is modeled as follows.

$$G'_A(t) = mG'_C(t) + c + \epsilon, \tag{6.25}$$

where m and c are the slope and intercept obtained from linear regression, and $\epsilon \sim N(0, \sigma^2)$ is zero-mean Gaussian noise. It is Gaussian because linear regression inherently minimizes the squared L2 norm of the fitting error, which in turn maximizes the likelihood that the errors were Gaussian. All three parameters were obtained from the training set.

We claim that $G'_A(t)$ in the test set is anomalous if the following condition holds:

$$|G'_A(t) - (mG'_C(t) + c)| > k\sigma, \tag{6.26}$$

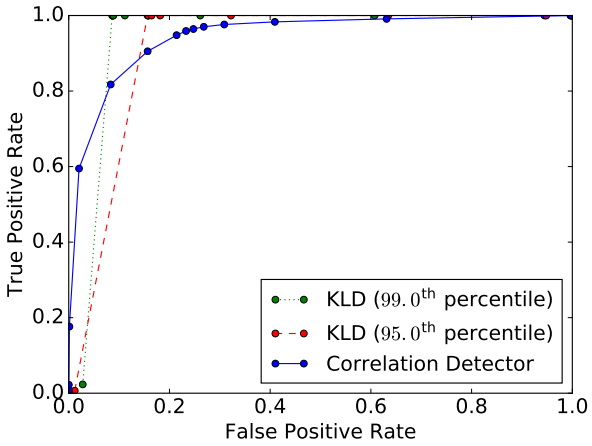
where m, c, and σ were obtained from the training set and G'_C was obtained at the same time as G'_A. k is the threshold parameter that determines the ROC for the correlation detector. The optimal attack against the correlation attack, which we refer to as the *correlation attack*, is achieved when A sets their generation reading as follows:

$$G'^*_A(t) = \min(mG'_C(t) + c + k\sigma, R_A(t)), \tag{6.27}$$

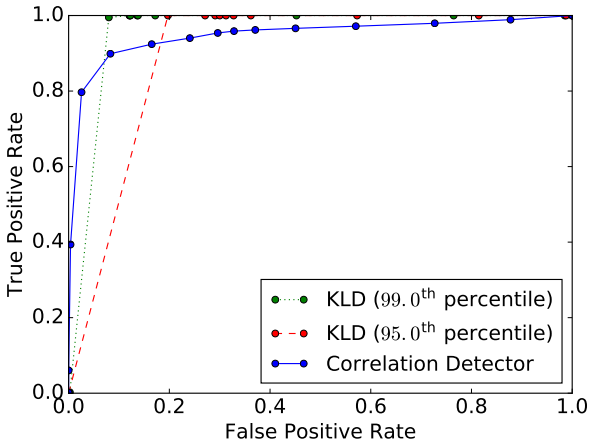
where $R_A(t)$ is A's rating, which should not be exceeded. In order to accomplish this attack, A would need to know that the utility is using C's readings for anomaly detection, and A would then need to monitor C's readings. In addition, A would need to know k. The difficulty of obtaining all that information may make this attack much less likely than the previous attacks, which looked only at A's own past readings.

We evaluate the KLD detector against the correlation detector in terms of how well they detect the *percentile attack*. The ROC curves for the results are presented in Fig. 6.10. It can be seen that the KLD detector narrowly outperforms the correlation detector for both the solar and wind datasets. In certain settings, as seen in Fig. 6.10(b), the correlation detector may achieve a more desirable trade-off with a FPR lower than that of the KLD detector.

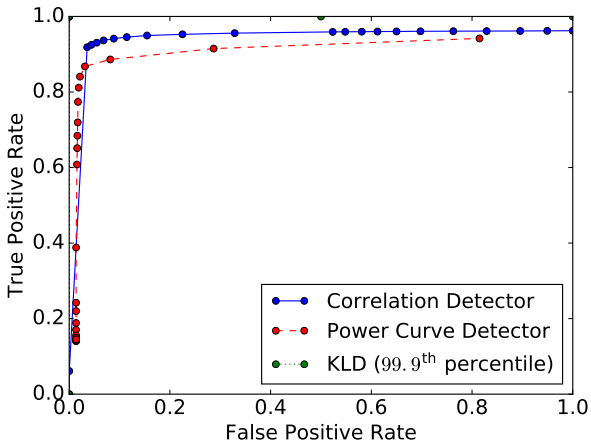
The KLD detector waits until a week of readings has been obtained and then determines that the readings have been over-reported consistently over the week. As a result, it produces



(a) Ausgrid residential-scale solar



(b) NREL utility-scale solar



(c) Engie wind turbines

Figure 6.10: ROC curves for DER mitigation methods.

a probability distribution of readings that differs greatly from the probability distribution of the training set. As seen in Fig. 6.10(c) for wind, the KLD detector achieves perfect detection performance, with an AUC of 1, at a 99.9^{th} percentile detection threshold. In the case of solar, its performance (in terms of AUC) is comparable with that of the correlation detector.

The fact that the correlation detector can work in real-time is an advantage over the KLD detector. However, we believe that an operator could use both detectors together, one in real-time and one at a periodicity of one week.

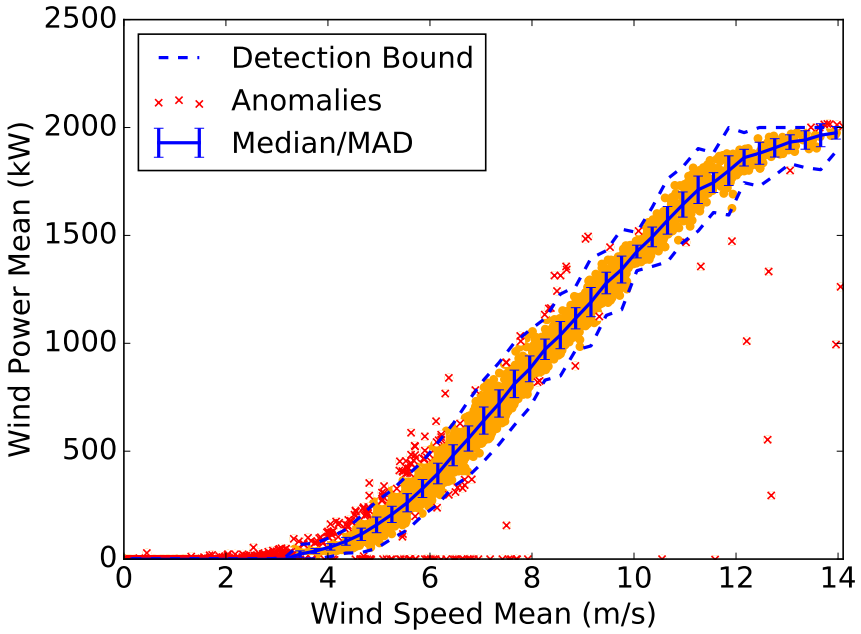


Figure 6.11: Statistical estimation of wind power from wind speed. The data points are classified into normal (yellow circles) and anomalies (red crosses).

6.4.4 Weather-Based Detectors

A limitation of using historical data in the previously described methods is that we need to assume that the training data have not been tampered with. If they have been tampered with, then statistical learning methods trained on the data could become biased in a way that ensures that the attacker escapes detection.

We now discuss the use of weather data to perform detection. The assumption is that weather data can be obtained from a completely different data source, which the attacker has not compromised. For example, a utility could use IBM's Deep Thunder [118], which provides wind speed and wind direction at turbine altitudes with a spatial resolution of 1 to 2 km. It also provides solar irradiance data at that spatial granularity.

For a fixed wind turbine configuration, the power produced can be obtained from wind speed measurements by using a well-known physical relationship called the *power curve*. Power curves for over 200 manufacturers' turbines are provided in [119] as look-up tables that map wind speed to expected power for each of those turbines. The operator could use those data to detect anomalous deviations between every wind turbine's expected and

reported output.

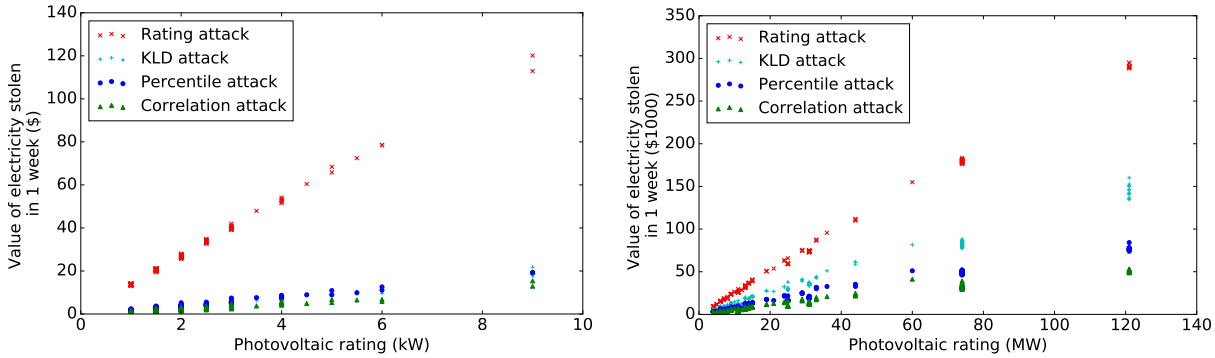
Similarly, solar generation can be predicted by using irradiance data along with solar panel configuration details such as the tilt angle (with respect to the axis perpendicular to the surface of the earth) and azimuth angle (with respect to true north). The NREL PVWatts calculator ([120]) estimates the expected solar output from the photovoltaic array. If the reported solar output significantly deviates from the expected output, the operator must investigate the cause of the deviation, as it may be indicative of an attack.

Since the Engie data contain wind power and wind speed measurements, we created an empirical model of the power curve, and used it to design what we call the *power curve detector*. The model is simple; we calculate the probability of wind power measurements, G'_A, given wind speed measurements, S'_A. We then extract distribution parameters for $P(G'_A|S'_A)$, such as the median and the median absolute deviation (MAD), and use those for anomaly detection as follows.

$$|G'_A(t) - \text{median}(G'_A|S'_A)| > k\text{MAD}(G'_A|S'_A), \tag{6.28}$$

where $P(G'_A|S'_A)$ is obtained from the training data. We used the medians and MADs because they are robust statistics, unlike the means and standard deviations. Figure 6.11 illustrates the power curve from the training set; the anomalies are not malicious, but are present in the dataset. If we had used means and standard deviations, those anomalies would have skewed the model and made it less effective for detection of anomalies in the test set. Once again, the optimal attack for this detector sets the generation at the detection threshold, while ensuring that it does not exceed the rating.

The ROC curve for the power curve detector is illustrated in Fig. 6.10(c) and is produced by varying k. For certain values of k, it produces a slightly lower false positive rate than the correlation detector, but overall it is inferior to the correlation detector. For $k = 0$ it does not have a high TPR or a high FPR because it turned out that the value of the percentile attack vector was lower than the median power for the given speeds.



(a) Residential-scale solar (Ausgrid)

(b) Utility-scale solar (NREL)

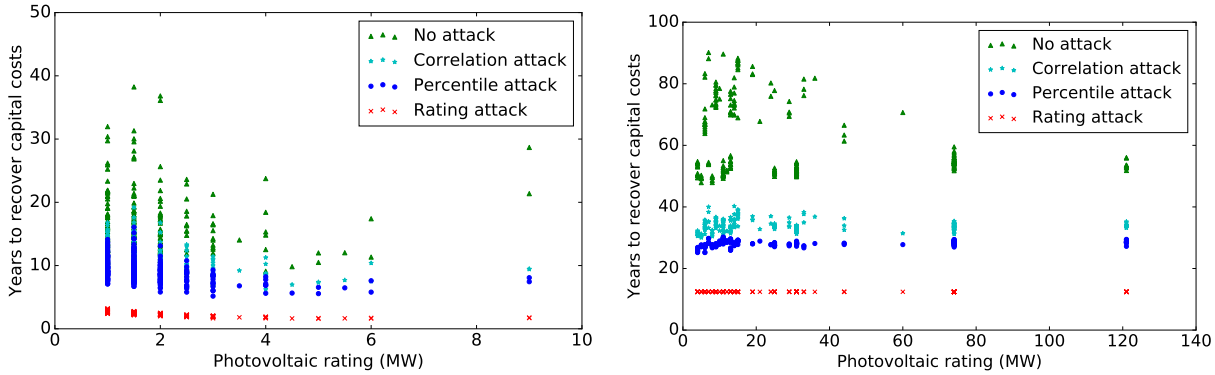
Figure 6.12: Solar data: value of electricity stolen through the optimal attacks that circumvent the rating and percentile-based detectors. The percentile-based detector mitigates the amount of electricity that can be stolen via the rating attack. Similarly, the correlation detector mitigates the percentile attack.

6.4.5 PCA-DBSCAN Detector

We had proposed this detector in the context of consumption readings, but we found that it was not suitable for DER fraud detection. In order for this detector to be successful, the generation patterns projected onto a lower-dimensional space would have to be tightly clustered so that density-based clustering could be used for anomaly detection. We found that that was not true of our solar and wind datasets because the data were not tightly clustered in the lower-dimensional space. Hence this detector is ineffective.

6.5 Profit Analysis of DER Fraud

The value of the electricity stolen for the solar scenarios is illustrated in Fig. 6.12, and is obtained by multiplying the quantity of electricity stolen by the electricity price. The Ausgrid feed-in tariff of $0.07/kWh [121] was applied to the Ausgrid dataset. The wholesale electricity price of $0.03/kWh was applied to the NREL dataset [122] and the Engie wind dataset [123]. The prices in California and in France are slightly higher, at approximately $0.05/kWh, and the price difference is attributed to operational costs [111], which do not contribute to profits. As seen in Fig. 6.12, the value of the electricity stolen in an average week varies linearly with the rating of the solar panel for all the optimal attacks.



(a) Residential-scale solar (Ausgrid)

(b) Utility-scale solar (NREL)

Figure 6.13: Time taken to recover capital costs of solar DER installations through different attack vectors. DERs are segregated based on their ratings.

Table 6.1: Average time (years) to recover DER capital costs

	Small-Scale Solar (Ausgrid)	**Utility-Scale Solar (NREL)**	**Utility-Scale Wind (NREL)**
No Attack	15.53	60.77	41.86
Rating Attack	2.44	12.43	7.61
Percentile Attack	9.14	28.03	9.42
Correlation Attack	10.68	33.79	20.50

Monetary gain is the attackers' primary incentive in the model described in Chapter 4. However, before a gain is made, they must recover their DER capital costs. In this section, we quantify how long it takes for them to recover those capital costs through the various attacks discussed in Section 6.4. Since capital costs are often large, the ability to speed up the return on investment through fraud may serve as an incentive for attackers to commit that fraud.

For the Ausgrid dataset, we use the solar installation costs given in [124]. Those costs can be corroborated with an independent source [125], and vary with the size of the installation.

For the NREL solar dataset, we obtained the cost per watt from the commercial solar power model given in [108]. We assume 1 MW panel array increments, and use the corresponding cost per watt of $2.03.

For the Engie wind dataset, we assume a cost of $4 million for the installation of each 2 MW turbine at each site. That is a conservative estimate taken from the range of $3–4 million given in [126].

We use the stolen electricity values described in Section 6.4 to calculate the time it would take to recover the capital costs of installing solar panels and wind farms. Figure 6.13 shows that time for the Ausgrid and NREL solar datasets. For each DER owner in each dataset, we plot the cost recovery time under the rating attack, percentile attack, correlation attack, and no attack. The owners are segregated by the ratings of their photovoltaics. For each rating, there can be considerable variance in the cost recovery time across DERs, as shown in Fig. 6.13. The reason is that the environmental conditions (exposure to sunlight) can differ across different DERs based on their location. In the Ausgrid dataset, the variance can be accounted for by the use of net metering, which takes the different consumption patterns into account.

The time to recover DER capital costs, averaged across all DERs in each dataset, is given in Table 6.1. "Small-scale solar" refers to residential-scale solar. For all three datasets, it is evident that, by committing fraud, attackers could recover their capital costs much faster than they could have if they had not committed fraud. That provides them with additional incentive for committing fraud. Across all types of DERs, the rating attack reduced the time it took to recover the capital costs by around 80% on average. The percentile attack benefited solar installations less than wind installations because solar generation was less erratic and approached the percentile-based threshold more often than wind generation approached that threshold. Therefore, there was less opportunity for fraud with the percentile attack with solar than with wind. As the correlation attack was the least beneficial to the attacker, the correlation detector was most beneficial to the defender. In that sense, the correlation detector mitigates the other attacks by forcing the attackers to wait much longer to recover their capital costs. The hope is that the additional wait time will disincentivize the attacker from committing fraud.

There is no closed-form optimal attack against the KLD detector, unlike the rating-based, percentile-based, and correlation detector. Assume that the attacker, A, knows the KLD detector parameters: the number of bins being used, B, and the percentile threshold cal-

culated on the training data, τ. The optimal attack vector for a week of readings, $\vec{G'_A}^*$, is

$$\vec{G'_A}^* = \arg\max_{\vec{G'_A}} \sum_{t=1}^{T} [\vec{G'_A}(t) - \vec{G_A}(t)] \tag{6.29}$$

$$\text{subject to } \mathrm{KLD}(\vec{G'_A}, \vec{G'}_{A,\text{training}}, B) \leq \tau, \tag{6.30}$$

where $\vec{G'}_{A,\text{training}}$ refers to the training data for A. The constraint is not a continuous function, and the space of readings is very large (countably finite because readings are rounded off to the nearest watt and bounded above by the rating of the DER). Therefore, maximization of profit on that space of attack vectors requires a search of an exponentially large space in T. The profit resulting from that optimal attack may not justify the associated cost of computational resources. We therefore recommend the KLD detector over the correlation detector.

6.6 Conclusion

In this chapter, we derived the optimal attack vectors for the PCA-DBSCAN detector and for the KLD detector in the context of consumption fraud. In doing so, we compared detectors based on the maximum gains that an attacker can make from each detector in the worst case. Since the worst-case attacker gains for the KLD detector exceeded those for the PCA-DBSCAN detector, the PCA-DBSCAN detector performed better. Although the KLD detector outperformed the PCA-DBSCAN detector against the integrated ARIMA attack, which was not optimal against either of the two detectors, the PCA-DBSCAN detector performed better against optimal attacks. Therefore, there is no clear winner between the PCA-DBSCAN detector and the KLD detector; the manner in which they are evaluated determines which one performs best. We do, however, believe that the KLD detector is better for most scenarios (if not the worst case, or optimal attack) because its underlying assumption about the data (that distributions of data points do not change drastically between weeks) is less restrictive than the assumption made in the PCA-DBSCAN model

(that time points in a week are approximately, linearly dependent).

In addition to evaluating optimal attacks in the context of consumption fraud, we presented and evaluated data-driven detection methods against DER fraud. We used ROC curves to quantify the TPRs and FPRs so that a utility can use a detector setting that has a suitable trade-off between the TPRs and FPRs. We used examples from wind and solar generation to illustrate how much an attacker would stand to gain monetarily from DER fraud. We showed that that gain can enable attackers to decrease the time it would take them to recover the capital cost of their solar or wind installations by over five times. We presented various detection mechanisms that could be used to detect and mitigate such fraud. The detectors were evaluated based on how much an attacker could possibly gain by evading them. That gain could be translated into the time it would take for the attacker to recover her installation costs. The evaluation was driven by freely available data from Australia (provided by Ausgrid), France (provided by Engie), and the U.S. (provided by NREL).

In conclusion, we demonstrated the claim in our thesis statement in the context of improving detection of generated fraud. Our empirical correlation detector mitigated the optimal attack rating against the rating-based detector by over 77%. In doing so, it increased the time it would take for an attacker to recover installation costs by 2.5 times that of the rating-based detector. The work in this chapter was peer-reviewed and published in [33].

CHAPTER 7

CYBER-ATTACKS ON PRIMARY FREQUENCY RESPONSE MECHANISMS IN GENERATORS

"War is 90% information."

– Napoleon Bonaparte

In this chapter, we address the ugly consequence of smart grids, which is that increased connectivity and access to information have made it possible for cyber-adversaries to remotely cause damage and outages to power grid equipment. In doing so, we discuss cyber-resilience, the final theme of the dissertation.

As part of the Aurora generator test in 2007, researchers at Idaho National Laboratories demonstrated that supervisory control and data-acquisition (SCADA) systems in generation controls can be compromised by a remote adversary [11]. Stuxnet was malware that was used to target centrifuges in a uranium enrichment facility in Iran in 2009, causing them to rotate beyond their safe limits and leading to physical damage. It is considered to be the first cyber-weapon that had an impact on the physical world because it could connect to programmable logic controllers that caused damage to the centrifuges by operating them beyond their safety limits [12]. The second such cyber-weapon, called Crash Override, created an outage in Ukraine in 2015 through crafting of malicious commands that could cause grid relays to open, disconnecting loads from generators [13]. Those cyber-weapons have set a dangerous precedent and have motivated research on cyber-resilience for energy-delivery systems. In this chapter, we study a new attack vector that needs to be protected against in order to prevent cyber-induced power outages. Such outages could undermine critical defense infrastructure and much of the economy, and could place the health and safety of millions of people at risk.

Outages result when loads are disconnected from generators by protective equipment called *relays*, which isolate faults in a power grid. Faults typically cause the grid frequency to change drastically from its nominal value, which is 60 Hz in the U.S. If the frequency drops below a

specified threshold, the relay disconnects the load so that a fault in one part of the grid does not have cascading effects on a different part of the grid. That disconnection is called *under-frequency load shedding* (*UFLS*), and it can happen in the absence of attacks. Load shedding could also be caused by malware, such as Crash Override, that exploits vulnerabilities in specific relay software to cause the disconnection of loads. From the perspective of the electric transmission network, the load may represent a small town, and when UFLS disconnects that load, it causes a regional outage. As a result, the total load connected to the generators is reduced, and the generators can be fully utilized to serve other loads in the grid (other towns) despite the disruption to the town in which the fault occurred.

There are frequency response mechanisms in power grids that provide resilience by restoring the grid frequency to prevent UFLS. In most parts of the world, synchronous generators, such as hydro and thermal plants, are called on to provide *frequency response* by adjusting generation in order to restore the grid frequency in the event of an excursion. The increase in renewable adoption has motivated system operators in many countries to consider using wind farms for frequency response [127–130]. Wind power is playing a major role in meeting electricity demand in an increasing number of countries, including Denmark (42% of demand in 2015), Germany (more than 60% in four states), and Uruguay (15.5%) [16]. European countries such as Ireland, Spain, and Germany already have protocol requirements in place for communications between system operators and wind farms for frequency response [131, 132].

7.1 Summary of Contributions

In this chapter we present the first study of attacks that target the primary frequency response mechanism through malware similar to Stuxnet and Crash Override. We provide some background on frequency response and UFLS in Sections 7.2 and 7.3, respectively. We describe the system model and the threat model in Sections 7.4 and 7.5, respectively.

We evaluate different attack parameters and their impact on the grid frequency by using the PowerWorld Simulator. In doing so, we obtain the parameters that correspond to the minimum effort required for an attacker to effect UFLS. We perform those simulation studies

for synchronous generators in Section 7.6 and wind turbine generators in Section 7.7. We propose defense strategies against the attacks in Section 7.8. There are other attacks that could cause loss of resilience, and those are discussed in Section 7.9. We present related work in Section 7.10 and conclude in Section 7.11.

7.2 A Brief Review of Frequency Response

There are three main mechanisms in today's power grids that make the grids resilient to faults that cause the frequency to drop. They are primary, secondary, and tertiary frequency response [133]. Primary frequency response (PFR) is a decentralized approach wherein generators monitor the grid frequency at their location and perform droop control to restore the frequency to its nominal value. Droop control essentially controls the rotation speed of the generator turbine by controlling the flow of water (in the case of hydro turbines) or steam (in the case of steam turbines). PFR takes effect immediately after the frequency has dropped below the nominal value. Secondary frequency response uses a centralized mechanism at the dispatch center to dispatch multiple generators from multiple areas of the grid to aid in frequency response. That mechanism is called *automatic generation control* (AGC), and it takes 30 seconds to take effect after a fault. Tertiary frequency response uses reserve generators to compensate for any generation deficit and takes effect 10 to 30 minutes after the fault.

Outages induced by UFLS are the focus of this chapter and are highlighted in the flowchart in Fig. 7.1. First, there must be a fault that causes a loss of generation and causes the frequency to drop. If PFR fails to restore the frequency to the nominal value, AGC is called on to do the job. If AGC fails, then tertiary frequency response is called on (which is not illustrated in Fig. 7.1). Despite those layers of protection, UFLS can result if AGC or tertiary frequency response kicks in too late. This chapter is primarily concerned with PFR, whose responsibility is to ensure that the grid frequency is safely restored (to a value above the stipulated threshold for UFLS) before AGC takes effect. If PFR fails to do so, then UFLS would occur and cause regional outages before secondary and tertiary frequency response have the opportunity to restore the frequency.

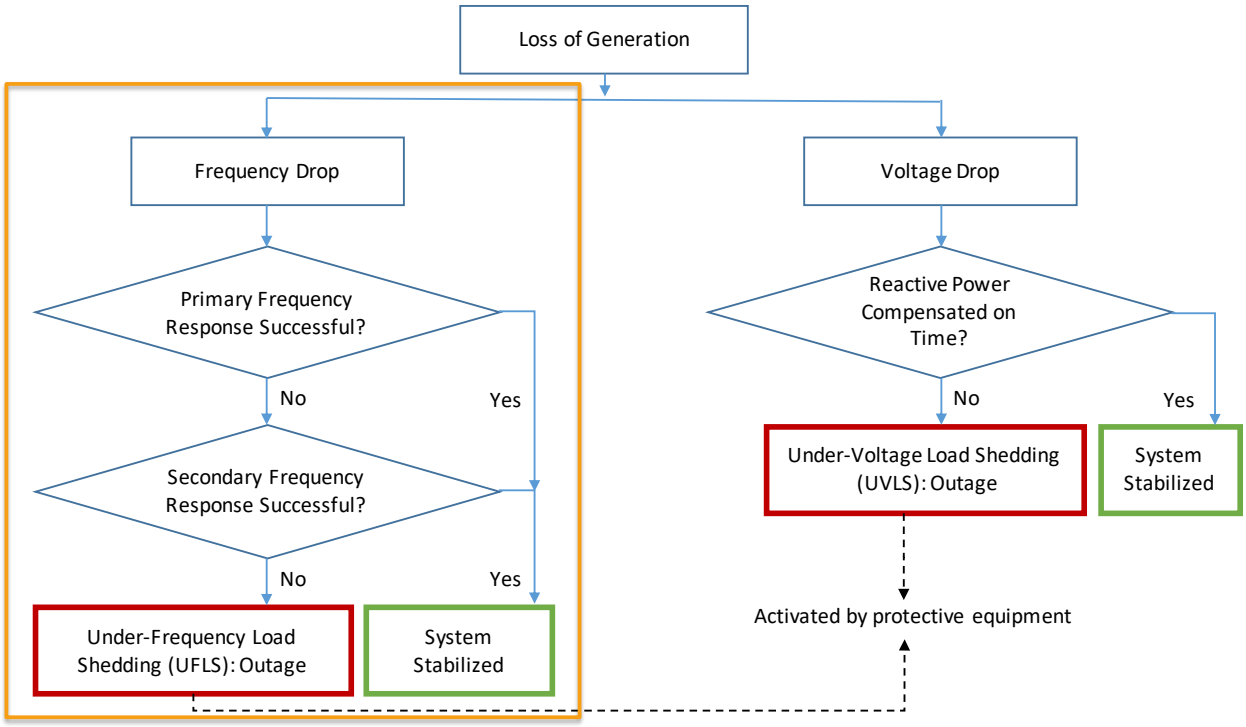


Figure 7.1: Flowchart depicting steps that lead to an outage or system stability after a loss of generation.

Note that outages can also result as a consequence of under-voltage load shedding, which is illustrated but not highlighted in Fig. 7.1. The voltage drops after the frequency drops because of a lack of reactive power in the network. There are many devices in the power grid, such as variable shunt reactors, synchronous condensers, and static VAR compensators that generate or absorb reactive power as needed to maintain the voltage at different parts of the grid. We assume that those devices are properly functioning so that under-voltage load shedding is not a concern.

This work is the first to explore attacks on PFR. Turbine controllers that provide PFR have been analog and air-gapped from computer networks until recent years, and those air-gaps protected against cyber-attacks on PFR. However, those controls have become increasingly connected to distributed control systems (DCS) in generator operational technology (OT) networks. Therefore, this work is timely in that it provides prevention, detection, and response strategies for new vulnerabilities that may have emerged as a consequence of technology modernization and increased connectivity.

7.3 Under-Frequency Load Shedding Threshold

Different countries stipulate different thresholds for the frequency under which UFLS is triggered. In North America, the North American Electric Reliability Council (NERC) stipulates the threshold in the PRC-006-1 standard [134] entitled "Automatic Under-frequency Load Shedding." According to that standard, after a major transient disturbance, UFLS will occur if the frequency falls below $0.575 \log(t) + 57.83$ Hz after a period of t seconds. The logarithmic threshold accounts for the tendency of the frequency to recover slowly over a period of time as a result of control mechanisms that provide frequency response. Therefore, the threshold is initially lower (58 Hz for the first two seconds) and later higher (58.68 Hz after 30 seconds). If the frequency remains below 58.68 Hz at 30 seconds, UFLS will be triggered, resulting in regional outages even before AGC can take effect.

7.4 System Model

Our system under study is the smart grid. In this section, we provide background on power grid analysis, generation controls, and network architectures in generation control centers.

7.4.1 Power Grid Analysis

The electric power transmission system comprises a multitude of generators and loads. The generators and loads are connected directly to buses, and the buses are connected by the transmission lines, transformers, and protective equipment. Broadly speaking, the power grid can be analyzed in two states: steady state and transient state. In steady-state analysis, an optimal power flow algorithm determines how much generation is dispatched to economically meet the load requirements over periods of time ranging from minutes to hours. Voltages, currents, and phase angles are calculated during the steady-state estimation process. Frequency is not considered in steady state because the system is assumed to be in a stable condition (i.e., no major frequency excursions). Transients, on the other hand, deal with unexpected failures that cause frequency excursions. Such failures can include short-

circuit faults, sudden tripping, or connection of large loads, sudden opening of transmission lines, among others. In this chapter, we are only interested in failures that result in the disconnection of one or more generators.

7.4.2 Turbine Controls

Most electricity generators are powered by turbines. The turbines create a rotating magnetic field, which in turn produces alternating current. The turbines are driven by the movement of fluids, such as steam, hydro, and wind. The flow of those fluids can be controlled by manipulating valves (steam), dams (hydro), or blade angles (wind). For the same turbine movement resistance, allowing greater flow produces greater rotation speed and thus increases the frequency of the power generated. When a fault occurs in the grid, causing the disconnection of one or more generators, all connected generators' turbines experience increased resistance to movement, and the flow into the turbines needs to be increased in order to maintain the required frequency. As an analogy, if a chariot drawn by four horses were to lose one horse, the remaining three horses would feel a larger burden and need to exert greater effort in order to maintain the same chariot speed. This chapter is focused on steam turbine generators, which are by far the most common type. The steam in steam turbine generators is produced from heat from burning fuels, such as coal and natural gas, or from nuclear fission.

To gain insight into the state of the practice, we interviewed David R. Brown, a senior consulting engineer for turbomachinery and generator control applications at Schneider Electric. He has been working on turbine controls for over 40 years, and we learned from him that generation control operators can now use DCS to modify the maximum power output of each generator by setting a parameter, which we refer to as P_{MAX}. That modification is accomplished by regulating the steam flow into the turbine via valves to ensure that the power output does not exceed P_{MAX}. The control of the flow of steam is key to providing frequency response, and that control has been analog and air-gapped until recent years. In recent years, according to Brown, the turbine controls have become increasingly automated with digital control and connectivity with DCS. The DCS provide connectivity between

multiple generation units (each containing a turbine) and a centralized control center at the generation facility. That network architecture allows remote control of turbine control settings through human-machine interfaces (HMI). Companies that provide digital control equipment for generator turbines include Voith, Woodward, Schneider Electric, ABB, and Honeywell.

7.4.3 Communication Network Design in Generation Centers

Apart from setting UFLS and related standards, NERC also provides a set of critical infrastructure protection (CIP) standards for device manufacturers and operators. Those standards specify best practices for ensuring security and resilience to cyber-attacks on industrial controls. Critical infrastructure control centers that comply with those standards, particularly with the electronic security perimeter standard (CIP-005-5), have their information technology (IT) and OT networks segmented as illustrated in Fig. 7.2.

As illustrated in Fig. 7.2, the IT network has direct access to the Internet through a perimeter firewall. That firewall provides basic IT security, restricting inbound connections and optionally performing intrusion detection. An additional firewall (or a virtual private network with access controls) exists to separate the IT and OT networks, restricting access to specific users who have privileges to access the OT network.

We assume that all generators comply with the NERC CIP standard. As a result, our baseline network security model reflects the current best practice.

7.5 Threat Model

In our threat model, we assume that the attacker has sufficient time and resources to plan and launch a sophisticated cyber-attack on power grid generators. Such an attacker could be a nation-state adversary, as was suspected for Stuxnet and Crash Override. The goal of the attacker is to create a regional outage by compromising PFR mechanisms in a subset of generators.

We assume that the attacker is unable to compromise the electric power transmission

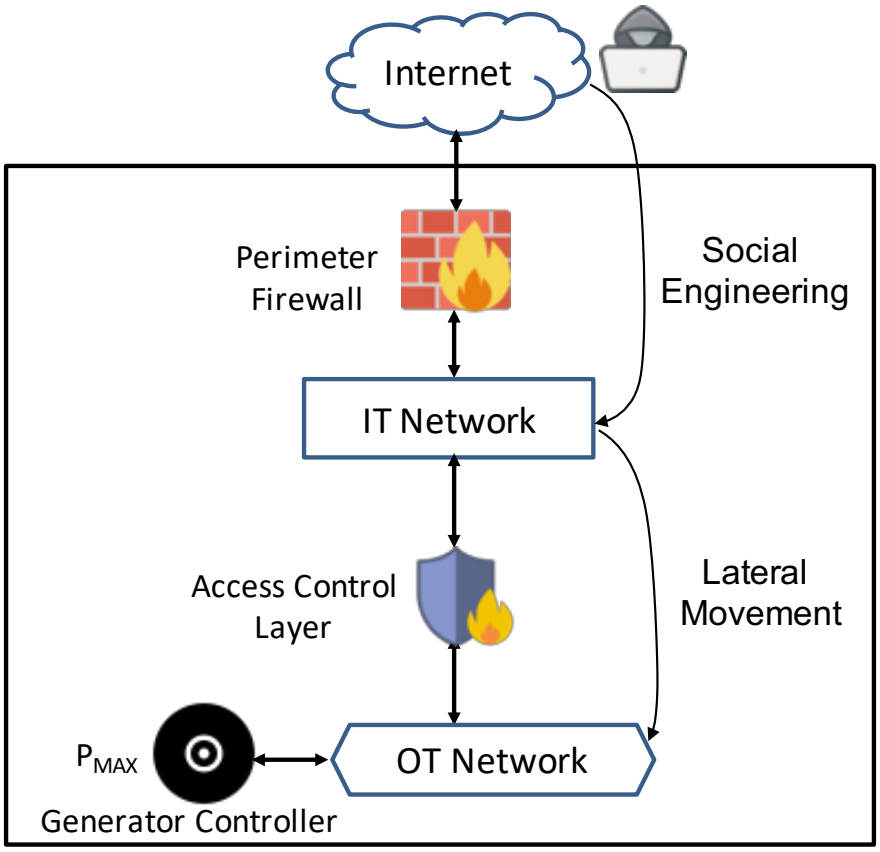


Figure 7.2: Attack path into the IT network, followed by the OT network for gaining access to the generator controls.

system operator facility (sometimes referred to as the *balancing authority*). Since those facilities have direct access to every sensitive piece of equipment in the power grid, gaining access to them would make it easy for an attacker to create an outage. As a result, we assume that great care has been taken to secure the system operator's physical and cyber territory.

We assume that in order to compromise generation facilities and cause an outage, the attacker would create an advanced persistent threat (APT) and take specific steps, as illustrated in Fig. 7.3. The logic bomb ensures that the malware remains latent and is triggered at a future time, when the system is heavily loaded and is more likely to suffer an outage after a failure (we will experimentally demonstrate this).

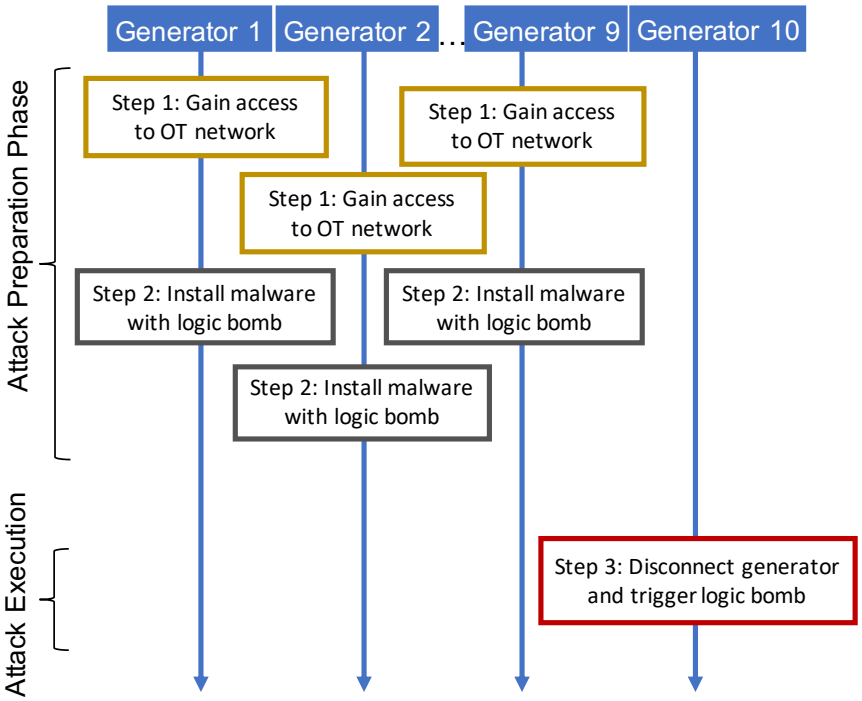


Figure 7.3: Steps taken by an attacker to cause an outage.

7.5.1 Gaining Access to the OT Network

Although the NERC CIP standards are detailed and well thought-out, they make no mention of social engineering attacks, which are used by over 80% of hackers, according to a 2016 survey [135]. Therefore, an attacker could gain access to the OT network by first entering the IT network through social engineering attacks, such as phishing, spear-phishing, and baiting. That is precisely how Crash Override entered the OT networks in Ukraine [13]. In order to gain access to the OT network from the IT network, the malware must gain access to credentials stored in the IT network that allow such access. Through those approaches, as illustrated in Fig. 7.2, the malware can circumvent firewalls and access control layers in the network. Once the malware has gained access to the OT network, it could send commands through the DCS to control the generation units.

7.5.2 Disconnection Attack

Once the attacker has gained access, they could simply instruct the DCS to disconnect all the generator units through generation circuit breakers. In order for that attack to cause an outage, the attacker would need to coordinate the disconnection of enough generators to create a UFLS scenario. In effect, the attacker would control a botnet, and synchronize the disconnection of multiple generators by communicating with the bot in each generator. While that attack might be effective, it would require management and coordination on the attacker's part. Also, it would not be subtle, and would raise suspicion of criminal intent.

At the time of this writing, there are very few generation circuit breakers that are connected to DCS systems. Therefore, the breakers may not be accessible to attackers, so disconnection attacks may not be feasible. One example of a network-connected generation circuit breaker control and monitoring system available at the time of this writing is the ABB GMS600 series [136].

7.5.3 PFR Restriction Attack

We discovered a different attack model, wherein the malware targets the PFR system by reducing P_{MAX} and effectively restricting PFR. It is essential that the malware enter multiple generators in order to trigger UFLS, as we shall demonstrate in our simulation study. To circumvent detection, the malware could be loaded with a logic bomb that would cause it to remain dormant and reduce P_{MAX} only on a specified day and time, by which time the attacker could ensure that the required number of generators have been compromised. The day and time could be chosen based on how much load is expected in the grid. Peak load times, for example, are the best-suited for causing UFLS through an attack. The reason is that any loss of generation would create a larger drop in frequency because of the greater load.

After reducing P_{MAX}, the malware could also compromise the inputs to the HMI at the generation control center to mask the attack and make it appear as though P_{MAX} has not been modified. Ultimately, as illustrated in Fig. 7.3, the attacker would need to cause a loss of generation after the logic bombs have been triggered in the malware. That could

be accomplished by malware as described previously in the context of the disconnection attack. The disconnection attack would require the disconnection of multiple generators to be effective. The PFR restriction attack could work even if only one generator is disconnected, and is therefore more subtle than the disconnection attack. Being subtle, the attack could be misattributed to mismanagement of generators, or non-malicious faults, throwing off suspicion of criminal intent.

In comparison to the disconnection attack, the PFR restriction attack takes less effort for the attacker to implement because no coordination is required. As long as the malware has been installed in the attack preparation phase on all the target generators before the fault is induced, UFLS could be effected. The reason is that generators are inherently synchronized to the grid frequency; the attacker does not need bots that communicate to facilitate synchronization. After one generator has been disconnected in the attack execution phase, the change in inertia will be felt at all connected generators as the frequency declines. The other generators will then try to perform PFR by increasing their generation to compensate for the loss of generation in the system, but that increase will be limited by the value of P_{MAX}. In our simulation study, we will show that UFLS could result as long as a large number of generators have been compromised and their P_{MAX} settings have been set to values that are substantially lower than the default.

7.6 Simulation Study for Synchronous Generators

We use the PowerWorld Simulator [137] because it can simulate both the steady-state and transient scenarios in a power grid. Also, PowerWorld comes with widely used simulation models that have been validated, requiring us to make only minimal parameter modifications in order to demonstrate attacks on PFR. Since such attacks cause frequency excursions, we need to be able to analyze transient behavior, and PowerWorld provides sufficient fidelity to do so.

Our power grid model is the standard IEEE 39-bus model of the New England grid, illustrated in Fig. 7.4; it has 10 generators and 46 lines. We used the standard IEEEG1 governor control [138] in which the control parameter P_{MAX} limits the maximum generation

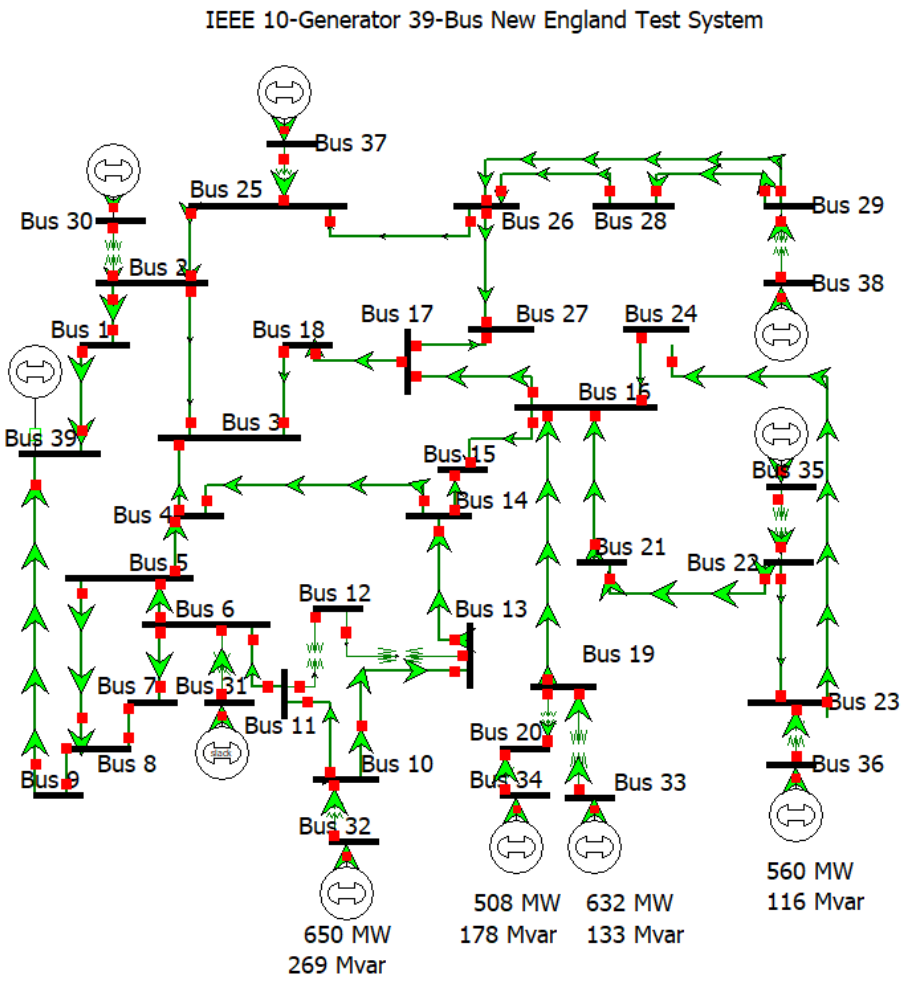


Figure 7.4: IEEE 10-generator 39-bus New England test system.

in the transient state, as discussed previously in the section on turbine controls. We modified the generator capacities from the base system to illustrate the attack at different levels of the total load relative to the total generation capacity in the system. We refer to the ratio of the total load (including transmission losses) to the total generation capacity as the *relative load* of the system. The greater the relative load, the more stressed the system would be before the transient event. In our threat model, the transient event corresponds to the disconnection of a generator in Step 3 of the attack steps illustrated in Fig. 7.3.

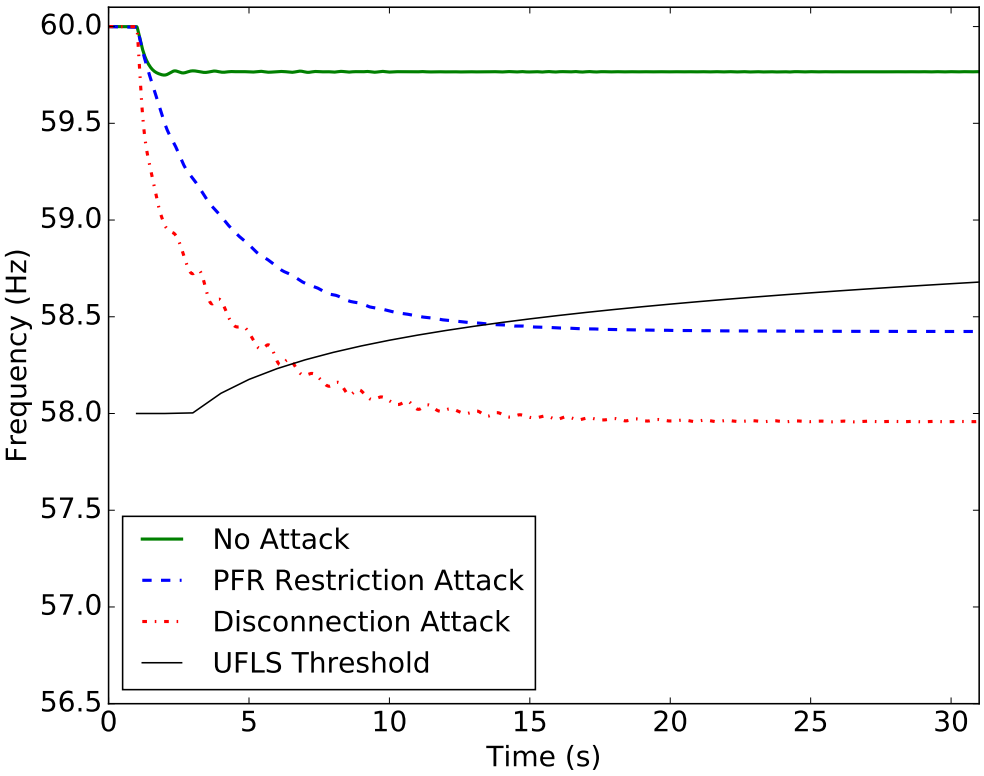


Figure 7.5: Frequency drop under different attacks after loss of generation.

7.6.1 Illustration of Impact of Attacks

Three scenarios from our evaluation are illustrated in Fig. 7.5. In all three scenarios, the default model was modified such that the relative load was 90%. We simulated a transient event in which the largest generator in the system (at Bus 39) was disconnected one second after the start of the simulation. As a result, the grid frequency started to decline from the nominal value of 60 Hz after one second. In the base case, PFR caused the frequency to stabilize at 59.77 Hz, which is a safe state above the UFLS threshold. That illustrates the inherent resilience of the grid to an attack that would cause the disconnection of the largest generator. As shown by the power flow arrows in Fig. 7.4, the other generators compensated for the loss of that generator by providing additional power through PFR.

For the disconnection attack, two generators in addition to the one at Bus 39 needed to be simultaneously disconnected in order to cause UFLS. The PFR restriction attack illustrated in Fig. 7.5 sets the P_{MAX} value to 90% of the generation capacity on 8 generators; no

Table 7.1: Disconnection attack parameters

Relative Load	Minimum Number of Generators Disconnected	Bus IDs of Generators
85%	4	39, 38, 32, 35
90%	3	39, 38, 32
95%	3	39, 38, 32
100%	2	39, 38

generator other than the one at Bus 39 was disconnected. In this experiment, the frequency fell below the UFLS threshold after 30 seconds, causing an outage right before AGC could take effect. Note that we illustrate the lower bound of the attacker's effort for all results presented in this section. Compromising more generators or reducing P_{MAX} to a greater extent would also cause UFLS, but that would require more effort from the attacker and possibly increase the risk of detection.

7.6.2 Evaluation of Disconnection Attack

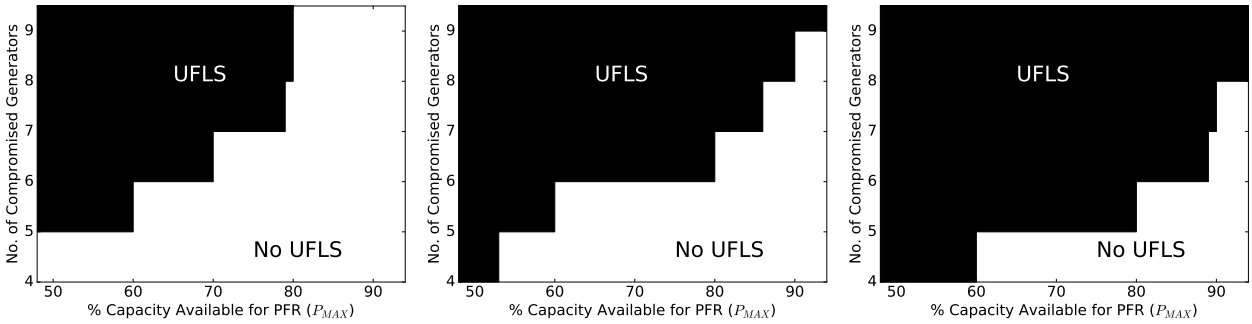
The disconnection attack has only one parameter, which is the number of generators to be disconnected. We evaluated the attack by identifying the minimum number of generators that would need to be disconnected for different relative load levels.

We used the IEEE 39-bus system shown in Fig. 7.4 to evaluate the disconnection attack. We programmed a transient event that would trigger the disconnection of generators in decreasing order of generation capacity. The attack was evaluated for different relative load settings, and the results are summarized in Table 7.1. As expected, the lower the relative load, the more capacity the system has for resilience against loss of generation. Therefore, it required the disconnection of four generators to create UFLS at a relative load of 85%. Conversely, when the load was 100%, only two generators needed to be disconnected in order to cause UFLS. The generators that were disconnected for each experiment are identified in Table 7.1 by the bus IDs to which they are connected. The case of 90% relative load corresponds to the curve for the disconnection attack in Fig. 7.5.

7.6.3 Evaluation of PFR Restriction Attack

We now explore the attack parameters of the PFR restriction attack. In all experiments, only the generator at Bus 39 was disconnected to induce a fault. The P_{MAX} values in other generators may have been compromised, but those generators were not disconnected.

We varied P_{MAX} as a percentage of each generator's capacity. In the base case (no attack), P_{MAX} was set to 120%, which is realistic because generators are allowed to exceed their capacity for a brief time period in order to enhance the system's resilience during a fault. If a generator were to operate beyond its rated capacity for too long, it could get damaged. For the attack cases, we modified P_{MAX}, varying it between 50% and 100% to determine settings that would result in UFLS. For each P_{MAX} and relative load setting, we used the simulator to obtain the minimum number of generators whose P_{MAX} values would need to be compromised in order to cause UFLS.



(a) Relative Load: 85% (b) Relative Load: 90% (c) Relative Load: 95%

Figure 7.6: Impact of attack parameters on UFLS.

We varied the relative load of the system by reducing the capacity of each generator proportionately. No other parameters (such as excitation) were modified. The results are summarized in Fig. 7.6. To help the reader interpret the plot, we describe two data points in Fig. 7.6(b), in which the relative load was set at 90%. The region corresponding to attack parameters that cause UFLS is shaded in black. First, restricting PFR by setting $60\% \leq P_{MAX} < 80\%$ required that at least six generators be compromised to cause UFLS. Second, setting $P_{MAX} = 90\%$ restricted PFR less, so the system became more resilient and the attack had to compromise all nine connected generators to cause UFLS. In general,

fewer generators needed to be compromised when P_{MAX} was reduced to further restrict the generation capacity available for PFR.

We now describe the trend across different values of relative load. As expected, greater relative load caused the UFLS region to become larger because the system was closer to its generation capacity and was unable to tolerate relatively small reductions of P_{MAX}. The UFLS regions for smaller relative loads were fully contained in the UFLS regions for larger relative loads. That indicates that an attacker is more likely to induce UFLS if they define a logic bomb (in Step 2 of Fig. 7.3) to activate PFR restriction and induce the fault (in Step 3 of Fig. 7.3) when the system is expected to operate at high loading conditions.

7.7 Simulation Study for Wind Turbine Generators

We used PowerWorld to demonstrate the PFR attack in the context of wind turbines by using a model of the Western Electric Coordinating Council (WECC) grid. The Council coordinates electricity for two Canadian provinces, 14 western states of the U.S., and northern Baja California, Mexico. We used the WECC Type 3 Wind Turbine Generator Model [139], which was created by a task force led by the Electric Power Research Institute with inputs from power companies, such as ABB and Vestas.

Broadly speaking, the power grid can be analyzed in two states: steady state and transient state. In steady-state analysis, an *optimal power flow* algorithm determines how much generation is required to meet the load requirements over periods of time ranging from minutes to hours. Frequency is not considered in steady state because the system is assumed to be in a stable condition (i.e., no major frequency excursions). Transient state, on the other hand, deals with unexpected failures that cause frequency excursions. We use PowerWorld because it can simulate both the transient and steady states of a power grid with sufficient fidelity (with millisecond resolution). Our power grid model is a widely used and validated approximation of the WECC power grid that is simplified to contain 9 logical buses and 3 logical generators [140]. The CAISO study by GE in [129] models the same system. Note that the logical buses and generators are scaled-down representations of the physical buses and generators.

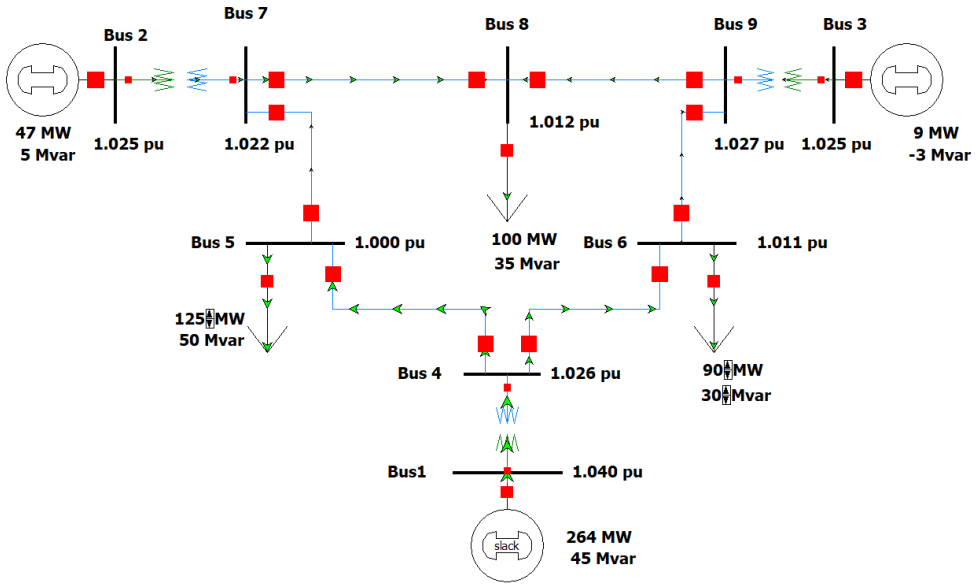


Figure 7.7: WECC 9-bus test system with steady-state generations and power flows.

We modified generators and loads slightly from the default WECC 9-bus test system; such modifications are a common practice in using any standard bus system for demonstrating specific scenarios. The single line diagram (with loads and steady-state generations) is shown in Fig. 7.7. The generators on buses 1 and 3 (henceforth referred to as *Gen 1* and *Gen 3*, respectively) represent several conventional steam turbine generation units. We modified the generator at bus 2 (Gen 2) to use a wind turbine generator (WTG) based on the WECC model. The reason for modifying the model was to match the scenario in related work by GE [129] so that a direct comparison of results could be made. In GE's scenario, a sudden loss of generation, amounting to 3% of the total steady-state generation, was simulated, and the response of the system to that sudden loss was studied. GE showed in [129] that the WECC is inherently resilient to such loss of power because of frequency response mechanisms. They argue that the loss of 3% of steady-state generation is a realistic scenario for a non-malicious fault, and is equivalent to losing a nuclear plant. In our 9-bus test system, we simulated that loss of generation by tripping Gen 3 (opening its breaker).

We kept the generation capacities constant in all experiments. Gen 2 and Gen 3 were configured to be "PV buses," allowing us to specify their steady-state contributions, which were lower than their capacities. Gen 1 picked up the slack when Gen 2 and Gen 3 could

not meet the demand.

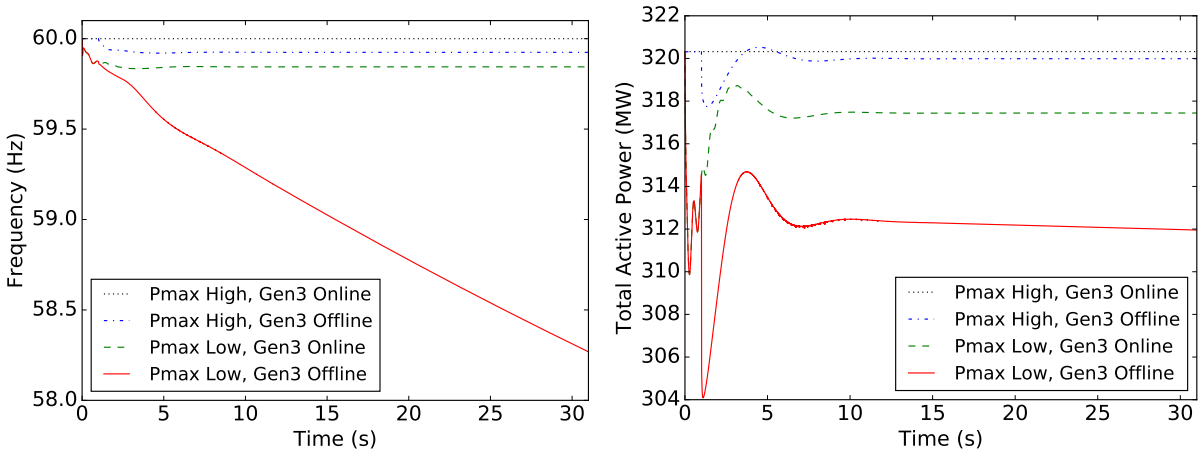
7.7.1 Attack Illustration

The three logical generators in the 9-bus test system play unique roles in our attack simulations. Gen 3 is the generator that is tripped, creating a transient event (fault) that causes the frequency to drop. Gen 1 and Gen 2 compensate for the loss of generation from Gen 3 and restore the frequency through PFR mechanisms within 30 seconds of the fault.

The malware in our attacker model targets the control system of Gen 2 and throttles its ability to respond to the loss of Gen 3. It does so by reducing the value of a parameter in the WTG model called P_{MAX}, which limits the active power output of the WTG. The model comprises a complex control loop, whose description is given in [139]. In that model, P_{MAX} serves to limit the amount of wind energy that is converted to electrical energy. The notion of P_{MAX} is not unique to the WECC Type 3 WTG model in [139]; it also exists in steam turbine control models, such as the IEEE G1 governor control [138]. Reducing P_{MAX} effectively limits PFR. In practice, P_{MAX} can be remotely set through the OT network through the use of a distributed control system in the wind farm.

It is essential that the malware enter multiple WTGs in order to trigger UFLS. In all our attack simulations, we assumed that all physical WTGs that constitute Gen 2 had been compromised by the attacker. To circumvent detection, the malware could be loaded with a logic bomb that would cause it to remain dormant and reduce P_{MAX} only on a specified day and time, by which time the attacker can ensure that the required number of generators have been compromised. The day and time can be chosen based on how much load is expected in the grid. Peak load times, for example, are the best times to cause UFLS through an attack. The reason is that any loss of generation would create a larger drop in frequency because of the greater load.

P_{MAX} can take a continuous range of values from 0 MW to the WTG rating for Gen 2. For synchronous generators (like Gen 1), P_{MAX} can go slightly above the generator's rating for a short period of time, beyond which the generators would suffer physical degradation. We set the P_{MAX} for Gen 1 at 20% above its rating and left it at that for all experiments described



(a) Average Load Bus Frequency vs. Time

(b) Total Active Power Generated vs. Time

Figure 7.8: Illustration of attack scenarios. For scenarios in which Gen 3 is offline, it was tripped at 1 sec.

in this section. For Gen 2 alone, we modified P_{MAX} as a consequence of the attack. The default setting, denoted by *high*, is the setting at the WTG rating (100% of its capacity); the compromised setting, denoted by *low*, restricts the output of the WTG (to 80% of its capacity) in transient events.

The impacts of four different attack settings are illustrated in Fig. 7.8. We compromised P_{MAX} at 0 seconds and tripped Gen 3 at 1 second. The effects were as follows:

1. P_{MAX} *High, Gen 3 Online*: In this scenario, there was no attack, and the frequency remained at the nominal value of 60 Hz.

2. P_{MAX} *High, Gen 3 Offline*: In this scenario, Gen 3 was tripped, but P_{MAX} was not compromised. The frequency was stabilized to a safe value slightly below 60 Hz. This illustrates the inherent resilience of the grid to loss of generation, and that resilience is due to PFR.

3. P_{MAX} *Low, Gen 3 Online*: In this scenario, P_{MAX} was compromised, but there was no loss of generation. Therefore, the system was not sufficiently stressed to cause UFLS.

4. P_{MAX} *Low, Gen 3 Offline*: In this scenario, P_{MAX} was compromised and then Gen 3 was tripped, and that led to UFLS. This scenario shows that limiting PFR by reducing

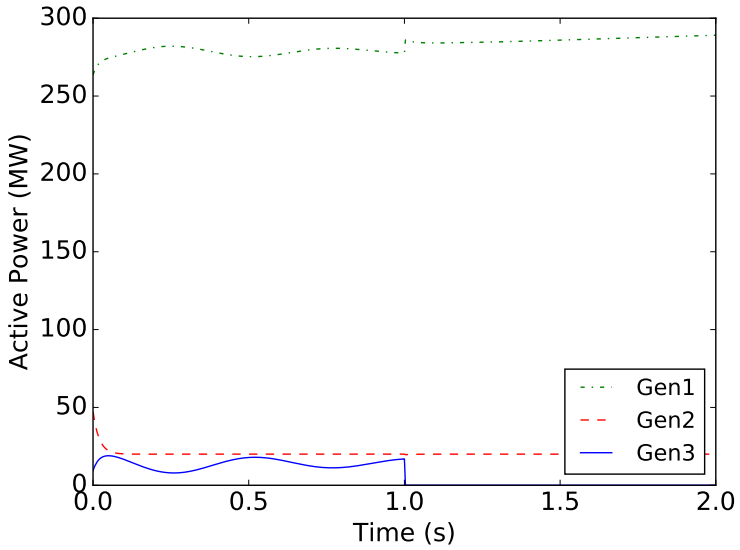


Figure 7.9: Generator outputs for P_{MAX} Low and Gen 3 tripped at 1 sec. Gen 2 does not participate in PFR, but Gen 1 does.

P_{MAX} undermines the resilience of the grid to the loss of Gen 3, and can lead to a blackout. The active power outputs in this scenario are illustrated in Fig. 7.9. According to PFR mechanisms (droop control), Gen 1 tries to compensate for lost power, but Gen 2 does not because it has been compromised.

7.7.2 Attack Evaluation

We varied the percentage of the load met by Gen 2, and evaluated the minimum change to P_{MAX} that needed to be made in order to create UFLS at 30 seconds after Gen 3 was tripped. In varying the contribution of Gen 2 to the load, we effectively evaluated different WTG penetration scenarios at steady state. Our scenarios can be directly compared with the scenario in the CAISO study by GE [129], in which the WTG penetration was 17%. While that study demonstrates resilience to loss of generation in the absence of attacks, our study shows how that resilience can be undermined by an attacker.

When P_{MAX} is reduced by a certain amount, the active power output is proportionally reduced. Figure 7.10 illustrates that the required reduction of P_{MAX} to cause UFLS changes with increase in penetration of wind turbine generation. P_{MAX} needs to be further decreased (more intense attack on generator) with the increase in the WTG penetration level. That

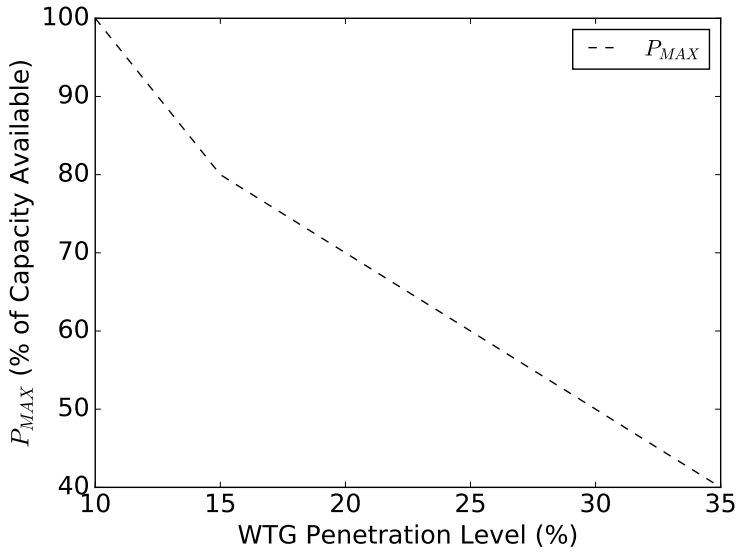


Figure 7.10: P_{MAX} setting required for causing UFLS, as a function of WTG penetration.

may seem counterintuitive because it implies that the attackers would need to make a more egregious (or easy to detect) compromise if they had access to a larger fraction of the steady-state WTG power to begin with. In other words, *increasing the WTG penetration inherently makes it harder for the attacker to avoid detection.* As seen in Fig. 7.10, an attacker needs to reduce power by nearly 80% (from 100%) to create a blackout when Gen 2 (the WTG) contributes 35% of the load. That is an encouraging result because it promotes increased wind integration for security.

The aforementioned counterintuitive behavior is explained as follows. When the percentage contribution of Gen 2 to the load increases, and Gen 3 is kept constant, the percentage contribution of Gen 1 must decrease. However, the capacity of Gen 1 was left unchanged, so Gen 1 was operating increasingly far below its potential as Gen 2's contribution increased. As a result, Gen 1 had more leeway to compensate for the loss of Gen 3. When the WTG penetration level was at 10%, however, Gen 1 was already operating close to its capacity, so it could not compensate as much for the loss of Gen 3. Hence, increasing WTG penetration while keeping the conventional generation capacity constant led to an improvement in security.

7.7.3 Detection and Mitigation

After reducing P_{MAX}, the malware could also compromise the inputs to the human-machine interface (HMI) at the generation control center to mask the attack and make it appear as though P_{MAX} has not been modified. In such a scenario, an alternative detection approach is needed. In this section, we propose a data-driven approach to detect reductions in P_{MAX}.

Dataset and Assumptions

We used the Engie wind power dataset described in Section 6.3.3 for our detector evaluation. We assume that this dataset is free of maliciously compromised measurements, and use the data to understand normal consumption behavior. We also assume that wind speed data can be estimated using out-of-band measurements or weather forecasting engines like IBM's Deep Thunder [118], and that those measurements cannot be modified by the attacker.

Detection Threshold

In Section 6.4.4, we proposed the use of a power curve in detecting generation fraud, and we explore that method in this section to detect the compromise of P_{MAX}. The power curve maps the wind speed to wind power for a WTG; the power curve for one 2 MW turbine in the Engie dataset is illustrated in Fig 7.11.

Using the Engie dataset, we created an empirical model of the power curve, which calculates the probability of wind power generation, G, given wind speed, S. From Fig 7.11, it can be seen that when the wind speed is greater than 14 m/s, the turbine generates at its rated capacity. Therefore, the *rated speed* is 14 m/s. Figure 7.12 uses a histogram to plot a probability distribution of wind power given wind speed is greater than 14 m/s using the data from Fig. 7.11.

We will use the histogram in Fig. 7.12 to design a detector that can detect whether P_{MAX} has been reduced. On a training dataset of 720 days of data at a 10 minute time resolution, we found that $P(G < 92.7\%|S > 14) = 0.02$. That means that if we were to set a detection threshold at 92.7% when the wind speed is greater than the rated speed, then, on our

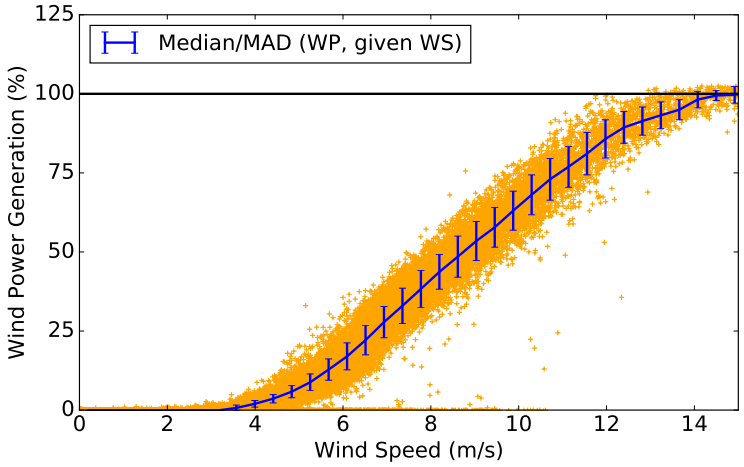


Figure 7.11: Power curve of one turbine in the Engie dataset used to model wind power, given wind speed. Wind power as expressed as a percentage of the rated capacity, which is 2 MW. The median and median absolute deviation (MAD) are illustrated for different wind speeds (divided into bins).

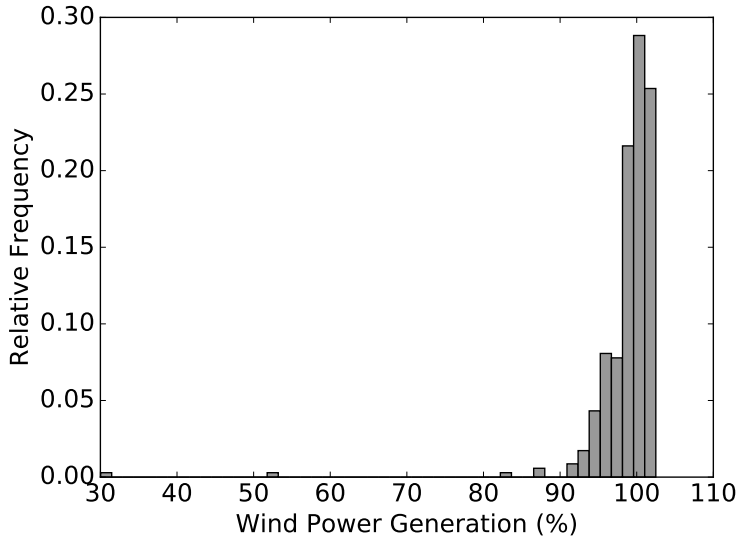


Figure 7.12: Conditional distribution of wind power generated by a 2 MW turbine in the Engie dataset, given that wind speed is greater than 14 m/s.

training dataset, the false positive rate would have been 2%. The threshold can be increased to obtain a smaller false positive rate, at the cost of a smaller detection rate. However, if P_{MAX} were reduced below 92.7% of the turbine's capacity, we would be able to detect the attack 100% of the time. This empirical detector is very effective because it would be very difficult for the attacker to cause UFLS by setting $P_{MAX} \geq 92.7\%$ (as seen in Fig. 7.10). We tested the false-positive rate on a hold-out set of 80 days of data at a 10 minute time resolution and found that it was 1.83%, which is comparable with the training false positive rate of 2%.

7.8 Defense Strategies

We now discuss prevention, detection, and response for the attacks presented in this chapter.

7.8.1 Prevention

We suggest that utilities not only comply with the NERC CIP standard, but also augment standard personnel training with additional training on social engineering attacks, such as phishing, spear-phishing, and baiting. Furthermore, e-mail filters in the IT network can be configured to block suspicious e-mails containing attachments or links to unknown Web domains. That would reduce the risk of downloading malware onto the IT network. In addition, we propose further segmentation of the IT network to ensure that personnel with access privileges to the OT network use a different, lower-privileged account when logging into computers for the purpose of browsing the Web and checking e-mail. That would ensure that any malware that might have been accidentally downloaded onto those computers would not have the privileges it needs to communicate with devices on the OT network.

If the logic bomb that is used to trigger the malware depends on remote control, the malware may need to communicate with the controller at an external IP address. Restricting outbound connections to unknown IP addresses would help prevent such remote control.

For generation controls in particular, additional measures can be taken. P_{MAX} is not altered on a frequent basis, so it can be made read-only through digital interfaces. If it does

ever need to be altered, the change can be done manually, as it was for decades until the recent advent of turbine control automation.

7.8.2 Detection

If an attacker were to modify P_{MAX} and also cause the HMI to show that no modification had taken place, it would be very difficult for the generation operator to detect the attack. We propose the use of an air-gapped measurement device such that there is no communication path through which the attacker can compromise that device. In particular, a tachometer can be installed to measure the number of rotations per minute (RPM), which is directly proportional to the grid frequency at the generator. That set-up is illustrated in Fig. 7.13. Any drop in the frequency resulting from a malicious reduction of P_{MAX} will be reflected in the tachometer reading, and alerts can be sent to technical staff if the frequency drops below preconfigured thresholds.

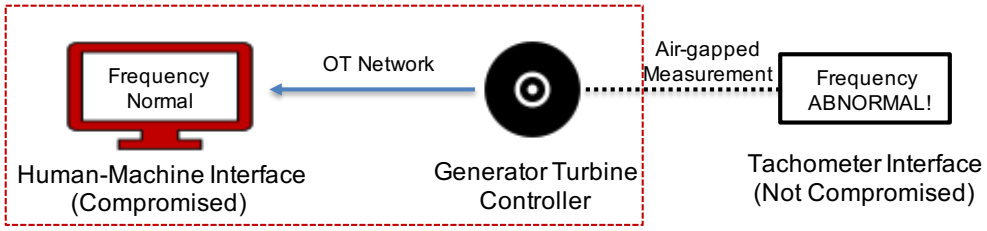
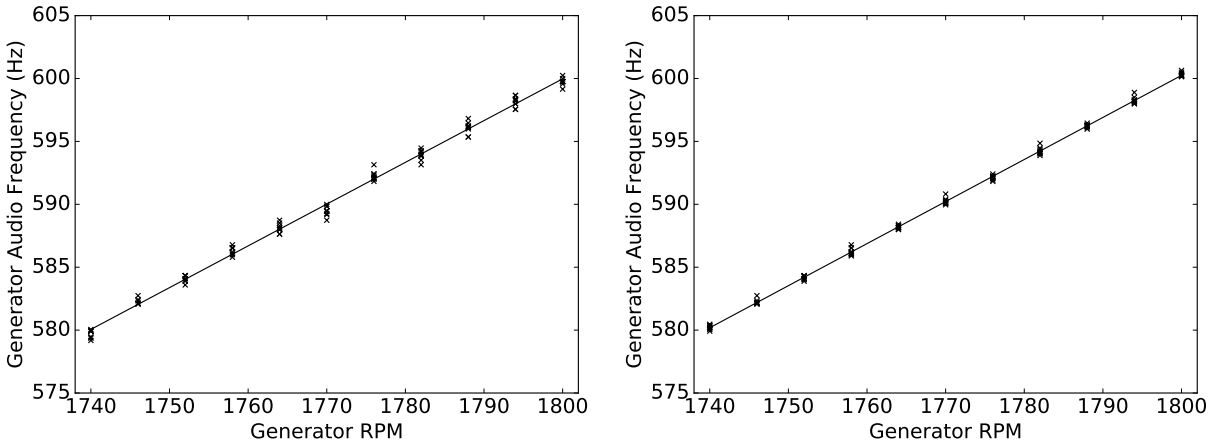


Figure 7.13: Out-of-band measurements from a tachometer can be used to validate frequency readings displayed on the HMI.

We discovered an additional detection method that leverages the correlation between the RPM of the generator and the sound that the generator makes. The author took the help of a fellow graduate student, Boya Hou, for measuring the frequency of the sound emanated by two small-scale generators in our university's power laboratory. Those were synchronous generators, manufactured by Advanced Motor Tech, and rated at 1.5 kW. The audio was captured using an iPhone and the time-domain audio signal was converted into the frequency domain using a fast Fourier transform. We picked the frequency corresponding to the highest amplitude and noted that as the audio frequency. For each of the generators, Boya separately measured the audio frequency at ten different RPMs, ten times each. That

produced a dataset of 100 readings per generator.

We used a scatter plot to observe the relationship between RPM and audio frequency of the two generators. As illustrated in Fig 7.14, that relationship is linear for both generators. Note that the relationship is slightly different for each generator, and we expect that difference to hold true for large generators as well. Using that linear relationship, a detection approach can be designed identical to the correlation detector given in Chapter 6 in Eqn. (6.26). That detection algorithm could be implemented in a smart phone app, and used by members of the generator operation staff to validate the RPM readings on the HMI. That approach assumes that the attacker does not have access to the operators' smart phones; the assumption is reasonable because the smart phone could be air-gapped from the generator IT and OT networks. We leave a more detailed exploration of this detection approach for future work.



(a) Generator 1

(b) Generator 2

Figure 7.14: Relationship between generator rotation speed (RPM) and the audio frequency of the sound emanated by the generator.

7.8.3 Response

If the system is believed to have been compromised, the operator can respond by overriding all controls to known safe settings and falling back to a manual operation mode wherein the controls are air-gapped from the OT network. The connection to the OT network can be restored after the malware has been removed or quarantined.

7.9 Additional Threat Models for Causing Power Outages

In this section, we briefly discuss additional threats to power grid resilience.

7.9.1 Attacks on AGC

Attacks on AGC are relevant only if UFLS is not triggered within 30 seconds. Even if an attacker were to compromise all the generators in one area of the power grid, AGC would allow for power to be fed into that area from other areas. However, that would work only if the areas are connected, through tie-lines, as they usually are in large-scale power grids. It would not work if the grid is isolated from other grids (as in a microgrid). AGC uses the area control error (ACE) to determine how much power needs to be delivered to the affected area (call it Area 1) from a healthy area (call it Area 2). An attacker could compromise the ACE reading to make it appear to the operator that there is no need for Area 2 to support Area 1. The ACE reading is a single point of failure, and compromising it would ensure that the generators in Area 2 will fail to receive the AGC signal from the operator to increase generation to compensate for the lost power in Area 1. PFR, on the other hand, is distributed, requiring the attacker to compromise multiple generators in order to create an outage. Because AGC is centralized, it is not only easier to attack, but also easier to defend because security mechanisms can be focused on maintaining the integrity of the single point of failure. PFR, on the other hand, requires that multiple, possibly heterogeneous turbine controllers be secured in order to prevent attacks.

7.9.2 Electricity Theft

Although the primary motivation for stealing electricity is monetary gain, the theft of large amounts of electricity could undermine the resilience of the grid. That was demonstrated in 2012 by the world's largest-ever blackout, during which 670 million people in India lost electric power largely because of consumption via theft that was unaccounted for in generation scheduling [141]. Theft detection using empirical models of generation and consumption data was presented in Chapters 4–6.

7.9.3 Attacks on Real-Time Pricing

We present a summary of our work on attacks on real-time pricing in Appendix C. Those attacks assume that consumers would automatically adapt their consumption patterns to changes in electricity prices. The mechanism for such adaptive consumption could be through automated demand response to prices. In that Appendix, we show that in theory, modifications to the real-time electricity price could cause consumption oscillations that could destabilize the grid and cause outages.

7.10 Related Work

Work on resilience of power grid control systems to attacks is presented in [142]. The authors of [143] discuss the compromise of control parameters related to frequency response in the power grid. Their attacks compromise secondary frequency response, or automatic generation control (AGC), while we compromised primary frequency response. Other papers that discuss attacks on AGC include [144–146].

Integrity attacks on power grid state estimation are presented in a number of papers, including [146–152]. Attacks on real-time pricing are presented in [88, 153, 154]. Those papers consider only the steady state of the power grid and therefore do not evaluate the inherent resilience of the grid to faults induced by attacks. Our work, and that of a few others, including [143,144,146], addresses that gap with an understanding of the real reason why outages happen: in response to severe frequency and voltage drops, both of which are transient states warranting transient analysis.

The authors of [155] claim to be the first to evaluate attacks on wind farms wherein the attacker is an outsider seeking to trip wind turbines and cause a large loss of generation. They perform a quantitative evaluation of attacks on a SCADA system within a wind farm by using a modified Bayesian attack graph model. However, they do not model the response of the power grid to the loss of generation. It is important to model that response because the grid may be resilient to the loss of generation, as we showed, and tripping turbines may not cause outages.

The use of wind power for frequency response is an active area of research, with recent industry implementations in countries including Spain, Germany, the U.K. [132], Ireland [131], and New Zealand [130]. The theory behind that use is described in [128] and [156]. A case study for California was presented in [129], and that is the most closely related work to ours except that they do not consider malicious compromise of generators. Frequency excursions with wind generation are simulated in [157] and [127], but the authors do not consider large excursions due to large loss of generation.

7.11 Conclusion

Primary frequency response is one of the most important mechanisms in today's power grids that make the grids resilient to faults, both malicious and non-malicious, that could lead to outages. In this chapter, we presented the first study of cyber-attacks that could cause power outages by infiltrating operational technology (OT) networks in generation control systems to inhibit primary frequency response.

We evaluated attacks that involve the malicious disconnection of generators and attacks that inhibit PFR. In both cases, the stronger the attack, the greater the risk of causing outages because of UFLS. We used PowerWorld to simulate the attacks and determine the weakest attack that can cause UFLS. In doing so, we obtained the range of attack parameters for which there is a risk of UFLS. By providing generation operators and technology providers with an understanding of the risks of such an attack, we hope that they will take concrete steps to mitigate that risk.

In the context of attacks on wind turbine generators, we showed that inhibiting wind power output can be detected using empirical models based on conditional probability distributions of wind power, given wind speed.

In conclusion, we demonstrated the claim in our thesis statement that empirical models can improve cyber-resilience in smart grids. In the context of cyber-attacks on wind turbines, we constructed an empirical model using wind power and wind speed data obtained from a wind farm in France. In the context of synchronous generators, we were unable to obtain real data, and instead relied on synthetic measurements from the PowerWorld simulator.

The work in this chapter is currently under peer-review as of this writing. Part of it has been accepted for publication and will appear in the November 2018 special issue of the IEEE Computer Magazine on resiliency in cyber-physical systems.

CHAPTER 8

CONCLUSION

"We can only see a short distance ahead, but we can see plenty there that needs to be done."

– Alan Turing

Smart grids have modernized traditional power grids through the use computer communication networks that enable data sharing between consumers, generators, and operators. The good consequence of that modernization is that it enables better utilization of clean energy resources. The bad consequence is that, if the data are not adequately protected, both consumers and generators can compromise the data for fraudulent monetary gains. The ugly consequence of increased connectivity is that it has become possible for cyber-attackers to remotely compromise power grid components, such as generator units, to cause power outages and disrupt daily life.

In this dissertation, we addressed the good, the bad, and the ugly consequences of smart grids through contributions made in three corresponding themes: resource utilization, fraud detection, and cyber-resilience. The problems in each theme are associated with costs that are in the billions of dollars annually. The contributions were organized in six different chapters, and are summarized as follows.

1. We enabled greater use of wind power by designing a data-driven approach that reduced the uncertainty associated with wind power predictions by over 20% in comparison to a heuristic model. The key innovation was in combining historic data with data from a fine-grained weather prediction engine to reduce prediction errors.

2. We improved utilization of demand response by developing an empirical model that can quantify the uncertainty associated with demand reduction in buildings. In comparison to a heuristic approach, the proposed statistical approach was able to increase demand reduction potential by three times.

3. We developed a framework for identifying and characterizing different ways by which meter fraud can be accomplished by both consumers and generators. In doing so, we identified seven attack classes under flat-rate and time-of-use pricing, of which five are new.

4. We devised empirical models to detect meter fraud for consumers by combining unsupervised learning methods in a novel way that was effective for anomaly detection. We evaluated the empirical models on detectors proposed in related work and showed that our detectors could reduce an attacker's monetary gains via fraud by 96%.

5. We derived the worst-case attacks against the detectors we proposed, and showed that realizing those attacks would be impractical. We showed that fraud committed by solar and wind power generators had the potential to reduce the time it could take for those generators to recover their capital generation costs by five times. Using an empirical detection method, we were able to mitigate generation fraud by 77%.

6. We studied attacks on synchronous and wind generation controls that could cause outages, and proposed ways to prevent such attacks. Using the PowerWorld simulator to evaluate our empirical detection method, we showed that attacks on wind power generators can be detected 100% of the time, with a 2% false-positive rate.

The first two of the aforementioned contributions can enable utilities to better utilize available options for wind power and demand reduction by assessing their associated risk or uncertainty; they can help save over a billion dollars per year of wasted clean energy potential. The next three contributions can help utilities all over the world mitigate fraud and save billions of dollars annually. The last contribution can prevent outages due to attacks on generation controls, and reduce the chance that the economy (which would suffer losses in billions of dollars) and daily livelihoods of people are disrupted because of a sudden loss of electricity. All six of the aforementioned contributions rely on the construction of suitable empirical models of generation and consumption.

In conclusion, we demonstrated our thesis statement: Empirical models of generation and consumption, constructed using machine learning and statistical methods, improve resource

utilization, fraud detection, and cyber-resilience in smart grids.

8.1 Future Directions

At the time of this writing, energy storage was prohibitively expensive at scale. While reservoirs have been used for large-scale energy storage, the elevated terrain requirements for reservoirs are available at few locations in the world. For that reason, storage was not a feasible solution to reduce wind power curtailments; that could be accomplished by storing excess wind power and using it at a later time. However, that may be a viable solution in the future if storage were to become economical at scale.

In the context of fraud detection, we proposed detectors and evaluated them independently. The detectors could potentially be combined using ensemble methods to combine the strengths of each detector. Furthermore, studies relating to multiple, simultaneous, fraudulent consumers could be conducted in future work.

In the context of cyber-resilience, we experimentally demonstrated the use out-of-band measurements for detecting attacks on wind turbine generators. We could not perform those evaluations for other types of generators because we did not have access to their data. However, once such access becomes available, that evaluation can be performed in the future.

8.2 Academic Recognition

The work in this dissertation was published in peer-reviewed journals and conference proceedings, as described in Appendix A. For the work in this dissertation, the author was selected to the Siebel Scholars Class of 2018 in energy sciences; the Siebel Scholars class recognizes the academic and leadership accomplishments of top graduate students from top universities around the world. The author was also awarded the Rambus Computer Engineering Fellowship in 2017, and an ECE Alumni Endowment Fellowship in 2018, both for excellence in research in computer engineering.

The work on fraud detection won seed-funding from the Siebel Energy Institute in 2015. The paper [89], titled "PCA-Based Method for Detecting Integrity Attacks on Advanced

Metering Infrastructure," won the Best Paper Award at the 2015 International conference on Quantitative Evaluation of SysTems (QEST). The paper [47], titled "ARIMA-Based Modeling and Validation of Consumption Readings in Power Grids," won the CIPRNet Young CRITIS Award (CYCA) at the 2015 International conference on Critical Information Infrastructure Security. That paper was also featured in the European Commissions' Critical Information Infrastructure Protection Newsletter.

8.3 Impact on Industry

The work that was presented in the dissertation was by nature predominantly applied, and some of it was transitioned to industry.

The work on wind power prediction (Chapter 2) was done in collaboration with IBM Research and Utopus Insights Inc. That work led to two U.S. patent applications, of which one was granted at the time of this writing. The proposed solution is being used by the Vermont Electric Power Company to help them better utilize their wind power generation capabilities.

At C3 IoT, the author worked on detecting consumers who had committed electricity theft among millions of consumers across Italy. In that role, the author used supervised learning methods for electricity theft detection, different from the unsupervised learning methods presented in Chapter 5. The availability of data corresponding to real attacks in practice made it possible for the author to use supervised learning methods at C3 IoT. The unavailability of that data for the research presented in this dissertation required the author to develop unsupervised learning methods and evaluate the detection methods based on their worst-case scenarios.

Apart from aforementioned industry engagements, the author interned at the ABB Corporate Research Center and contributed to ABB's energy management research efforts (related to Chapter 3). He also interned at Cisco Systems Inc. and contributed both research insights and production code to boost the cyber-resilience of their AMI networks.

APPENDIX A

PUBLICATIONS RELATED TO THE DISSERTATION

A.1 Peer-Reviewed Publications

A.1.1 Resource Utilization

1. Varun Badrinath Krishna, Wander S. Wadman, Younghun Kim, "NowCasting: Accurate and Precise Short-Term Wind Power Prediction using Hyperlocal Wind Forecasts," in *ACM Conference on Future Energy Systems (ACM E-Energy 2018)*. Karlsruhe, Germany, pp. 63–74. ACM, 2018.

2. Deokwoo Jung, Varun Badrinath Krishna, William Temple, David K. Y. Yau, "Data-Driven Evaluation of Building Demand Response Capacity," in *IEEE International Conference on Smart Grid Communications (IEEE SmartGridComm 2014)*, pp 547-553. IEEE, 2014.

3. Deokwoo Jung, Varun Badrinath Krishna, Ngo Quang Minh Khiem, Hoang Hai Nguyen, David K. Y. Yau, "EnergyTrack: Sensor-Driven Energy Use Analysis System," in *ACM Systems for Energy-Efficient Buildings (ACM BuildSys 2013)*. ACM, 2013. Best Paper Candidate.

A.1.2 Fraud Detection

1. Varun Badrinath Krishna, Carl A. Gunter, and William H. Sanders, "Mitigating Electricity Theft and DER Fraud," in *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Signal and Information Processing for Critical Infrastructures*, vol. 12, no 4, pp. 790-805. IEEE, August 2018.

2. Varun Badrinath Krishna, Kiryung Lee, Gabriel A. Weaver, Ravishankar K. Iyer, and William H. Sanders, "F-DETA: A Framework for Detecting Electricity Theft Attacks in Smart Grids," in *IEEE/IFIP International Conference on Dependable Systems and Networks (IEEE/IFIP DSN 2016)*, pp. 407-418, June 2016.

3. Varun Badrinath Krishna, Ravishankar K. Iyer, and William H. Sanders, "ARIMA-Based Modeling and Validation of Consumption Readings in Power Grids," in *10th International Conference on Critical Information Infrastructure Security (CRITIS 2015)*. Springer LNCS, 2015. CIPRNet Young CRITIS Award Winner (Best Paper by researcher under 32 years of age).

4. Varun Badrinath Krishna, Gabriel A. Weaver, and William H. Sanders, "PCA-Based Method for Detecting Integrity Attacks on Advanced Metering Infrastructure," in *12th International Conference on Quantitative Evaluation of Systems (QEST 2015)*. Springer LNCS Vol 9259. Best Paper Award Winner.

A.1.3 Cyber-Resilience

1. Varun Badrinath Krishna, Ziping Wu, Vaidehi Ambardekar, Richard Macwan, and William H. Sanders, "Cyber-Attacks on Primary Frequency Response Mechanisms in Power Grids," in *IEEE Computer: Special issue on Resiliency in Cyber-Physical Systems*. IEEE, 2018. In press.

2. Rui Tan, Varun Badrinath Krishna, David K. Y. Yau, Zbigniew Kalbarczyk, "Integrity Attacks on Real-Time Pricing in Electric Power Grids," in *ACM Transactions on Information and System Security (ACM TISSEC 2015)*, vol 18, no. 2, pp. 5:1-5:33. ACM, July 2015.

3. Rui Tan, Varun Badrinath Krishna, David K. Y. Yau, Zbigniew Kalbarczyk, "Impact of Integrity Attacks on Real-Time Pricing in Smart Grids," in *Proceedings of ACM Conference on Computer and Communications Security (ACM CCS'13)*. Berlin, Germany, pp. 439–450. ACM, 2013.

A.2 Patents

1. Varun Badrinath Krishna, Younghun Kim, Tarun Kumar, Wander S. Wadman, and Kevin W. Warren, "Reducing curtailment of wind power generation by improving wind power prediction accurac," US Patent US10041475B1 granted Aug 2018, filed Feb 2017.

2. Varun Badrinath Krishna, Younghun Kim, Tarun Kumar, Wander S. Wadman, and Kevin W. Warren, "Reducing curtailment of wind power generation by improving wind speed prediction accuracy," US Patent application 15/426,544 filed Feb 2017.

A.3 Posters and Demos

1. Varun Badrinath Krishna, Deokwoo Jung, Ngo Quang Minh Khiem, Hoang Hai Nguyen, and David K. Y. Yau, "DEMO Abstract- EnergyTrack: Sensor-Driven Energy Use Analysis System," in *ACM Systems for Energy-Efficient Buildings (ACM BuildSys 2013)*.

2. Juran Kirihara, Varun Badrinath Krishna and William H. Sanders, "Efficient Forecasting for Validating Smart Meter Measurements," in *IEEE Power and Energy Conference at Illinois (PECI 2016)*.

A.4 Newsletter Entry

- Varun Badrinath Krishna, "ARIMA-Based Modeling and Validation of Consumption Readings in Power Grids," in *European Critical Information Infrastructure Protection Newsletter*, vol 10. no 1.

APPENDIX B

EVALUATION OF PCA-DBSCAN AND KLD DETECTOR AGAINST THE INTEGRATED ARIMA ATTACK

In this section, we evaluate the four detectors proposed in this chapter against the integrated ARIMA attack. In doing so, we use the realizations of the integrated ARIMA attack illustrated in Fig. 5.8. We use three performance metrics described as follows.

Metric 1: The percentage of consumers for whom the detector successfully detected the attack. For Attack Class 1B, this directly translates to the percentage of neighbors who were protected from the attack by the detector.

Metric 2: The maximum amount of electricity stolen over a period of one week, as a result of the attacks' going undetected as per Metric 1. For Attack Class 1B, Metric 2 is the sum of the electricity stolen from all consumers for the period of one week. For Attack Classes 2A/2B, Metric 2 is the maximum amount of electricity that was stolen by a single attacker by under-reporting her own consumption. We include in Metric 2 the monetary profit associated with the electricity stolen. Recall that this profit can be attained by a realization of Attack Classes 3A/3B through leveraging of variable electricity prices, without stealing of electricity.

Metric 3: The area under the receiver operating characteristic (ROC) curve. This is a standard metric in detection theory that favors methods that achieve a high true-positive rate and a low false-positive rate.

In order to obtain the monetary gain for the attacker, we assume the following pricing system, which is based on the rates set by Electricity Ireland [158] (as our dataset comes from Ireland). The peak price is \$0.21/kWh (0.195€/kWh), and is valid from 9:00 A.M. to midnight. The off-peak price is \$0.18/kWh (0.172€/kWh), and is valid from midnight to 9:00 A.M. Adopting this TOU pricing scheme allows us to make a fair comparison between Attack Classes 1B–3B, as all of these classes work under TOU pricing (refer back to Table 4.1). Note

Table B.1: Results for Metric 1: Percentage of consumers for whom the detector successfully detected the attack

Electricity Theft Detector	**1B**	**2A/2B**	**3A/3B**
ARIMA detector	0%	0%	0%
Integrated ARIMA detector	0.6%	10.2%	0%
KLD detector (5% significance)	90.8%	64.4%	73.8%
KLD detector (10% significance)	94.0%	81.6%	82.0%

that the attack classes also work under RTP, but TOU is a far more widespread scheme than RTP, and data are available to make realistic assumptions about prices for TOU.

B.1 Results for Metrics 1 & 2

B.1.1 Results for Attack Class 1B

By design, the Integrated ARIMA attack evaded the ARIMA detector and the Integrated ARIMA detector. However, as shown in Table B.1, the KLD detector detected the attack for 90.8% and 94.0% of consumers at the 5% and 10% significance levels, respectively. Therefore, with the KLD detector, we provide a solution for detecting the integrated ARIMA attack, addressing the obvious gap in [47].

The electricity stolen because of the failure of the detectors is given in Table B.2. The electricity stolen as a result of evading the integrated ARIMA detector was less than that stolen as a result of evading the ARIMA detector by a factor of 78.1%. That quantifies the improvement of the Integrated ARIMA detector over the ARIMA detector in mitigating theft. After adding the KLD detector as an additional layer of detection, we observed an improvement of 96% over the integrated ARIMA detector for the worst-case attack. As a result, *the KLD detector almost completely mitigated theft through the integrated ARIMA attack (Attack Class 1B).* In addition, the KLD detector was found to be two orders of magnitude faster than the ARIMA methods proposed in prior work.

A 0% false-positive rate was observed in the test set for 29.2% of consumers when the KLD detector was set at the 5% significance level (and for 14.8% of consumers at the 10% significance level). The distribution of false-positive rates across multiple consumers

resembled an exponential distribution in which the largest fraction had a 0% false-positive rate.

B.1.2 Results for Attack Classes 2A/2B

We evaluated the four detectors against the integrated ARIMA attack as a realization of Attack Classes 2A/2B. The attack was supposed to circumvent (or be completely mitigated by) the integrated ARIMA detector. However, it was detected for 10.2% of consumers because the consumption readings were so low to begin with that the random numbers generated by the truncated normal distribution failed to maintain an average high enough to go undetected by the integrated ARIMA detector. The KLD detector performed much better, as shown in Table B.1. By design, the false-positive results were the same as for Attack Class 1B, since the test set and the detection approach were the same; only the attack vectors to test true positives were different.

Since the gains from Attack Classes 2A/2B are obtained by under-reporting the attacker's consumption, one may suspect that Mallory would stand to gain the most by being the largest consumer. However, the maximum amount of electricity stolen in a week via the ARIMA attack was achieved by the second-largest consumer in our dataset (Consumer 1330), and the corresponding value was $2,687$ kWh. The reason was that the lower bound on the confidence interval for Consumer 1330 was lower than it was for the largest consumer in our dataset (Consumer 1411). Thus, the readings for Consumer 1330 could be further under-reported.

The largest amount of electricity stolen in a week via the integrated ARIMA attack was achieved by the eleventh largest consumer in our dataset (Consumer 1333), and the corresponding value was $1,541$ kWh. In comparison, $1,382$ kWh was stolen from Consumer 1330 via the integrated ARIMA attack. Upon investigating this unexpected result, we found that it happened because the minimum of the average values for the weeks in the training set for Consumer 1333 was lower than it was for Consumer 1330. That implies that the mean of the truncated normal distribution for Consumer 1333 was correspondingly lower than it was for Consumer 1330, providing more room for Mallory to steal electricity.

At the 5% significance level, the integrated ARIMA attack by Consumer 1333 was not

Table B.2: Results for Metric 2: Maximum gains for attacker in one week as a result of circumventing theft detectors

Electricity Theft Detector	**Attack Class**	**1B**	**2A/2B**	**3A/3B**
ARIMA detector	Stolen (kWh)	367,777	2,687	0
	Profit ($)	72,797	542	14.3
Integrated ARIMA detector	Stolen (kWh)	80,535	1,541	0
	Profit ($)	15,647	297	14.3
KLD detector (5% significance)	Stolen (kWh)	3,575	1,541	0
	Profit ($)	694	297	14.3
PCA-DBSCAN detector	Stolen (kWh)	2,447	1,382	0
	Profit ($)	485	283	14.3
KLD detector (10% significance)	Stolen (kWh)	2,604	237	0
	Profit ($)	508	49	14.3

detected by the KLD detector in at least one of the 50 simulation trajectories. Hence we say that the detector failed for that attack, and the worst-case electricity stolen remains $1,541$ kWh. At the 10% significance level, however, the attack by Consumer 1333 was detected. The worst-case amount of electricity stolen despite the detector setting was only 237 kWh, which is 84.6% less than what it was under the integrated ARIMA detector (see Table B.2).

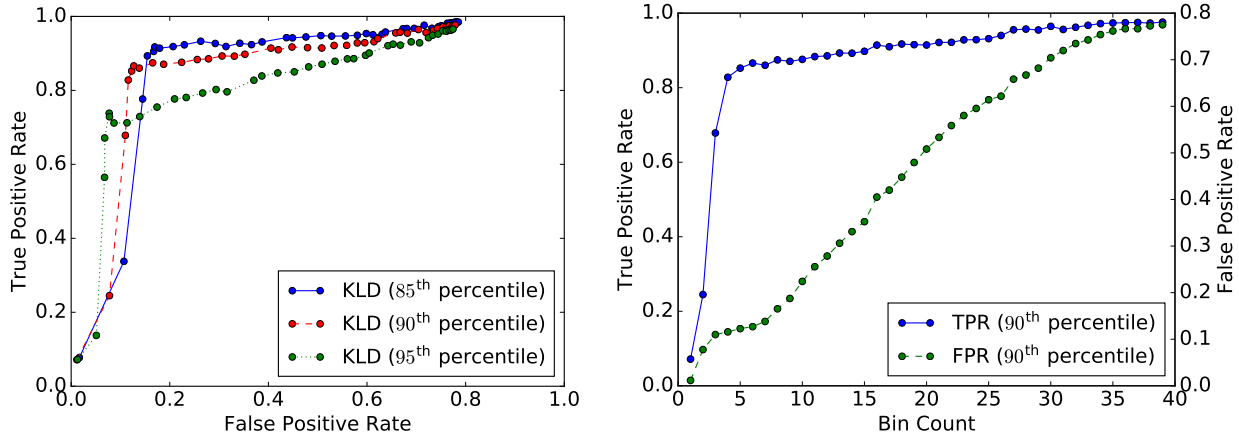
The lesson from these results is that *Mallory does not need to be the largest consumer in order to gain the most from theft.* It is also evident that the amount of electricity that can be stolen by circumventing the KLD detector via Attack Classes 2A/2B is an order of magnitude less than the amount for Attack Class 1B. This validates our earlier claim that Attack Class 1B is indeed the most advantageous attack class for Mallory.

B.1.3 Results for Attack Classes 3A/3B

All four detectors would fail to detect the Optimal Swap attack as it does not alter the distribution of consumption readings for the week. In order to detect the attack, the KLD detector needs to be modified by conditioning on the electricity price. Since we have two prices, we can split the X distribution into two distributions, one for peak period consumption readings and one for off-peak consumption readings. This idea can be extended from two distributions to multiple distributions, each conditioned on an electricity price in the case of RTP systems. We believe that if that method of conditioning is used, the KLD detec-

tor can also be used to detect Attack Class 4B. Conditioning also reduced the false-positive rate; a 0% false-positive rate was observed in the test set for 55% of consumers at the 10% significance level.

In the case of Attack Classes 3A/3B, the Optimal Swap attack was not detected by the ARIMA detector or integrated ARIMA detector, as shown in Table B.1. The KLD detector (conditioning on prices) was again successful in detecting this attack. The maximum monetary benefit was for Consumer 1254 (amounting to $14.3), whose Optimal Swap attack circumvented all four detectors by not significantly altering the distribution of consumption readings in the week. Even so, the benefit was so small that we believe Mallory would not invest the effort required to inject an instance of Attack Classes 3A/3B alone. She may, however, inject an attack that combines Attack Class 3B with Attack Classes 1B and/or 2B.



(a) KLD ROC Curve

(b) KLD TPR & FPR vs. B

Figure B.1: ROC for KLD detector on the integrated ARIMA attack. (a) ROC curves for three different thresholds on the KLD distribution. (b) TPRs and FPRs across different bin sizes (B) at a threshold set at the 90$^{\text{th}}$ percentile.

B.2 ROC for the KLD Detector

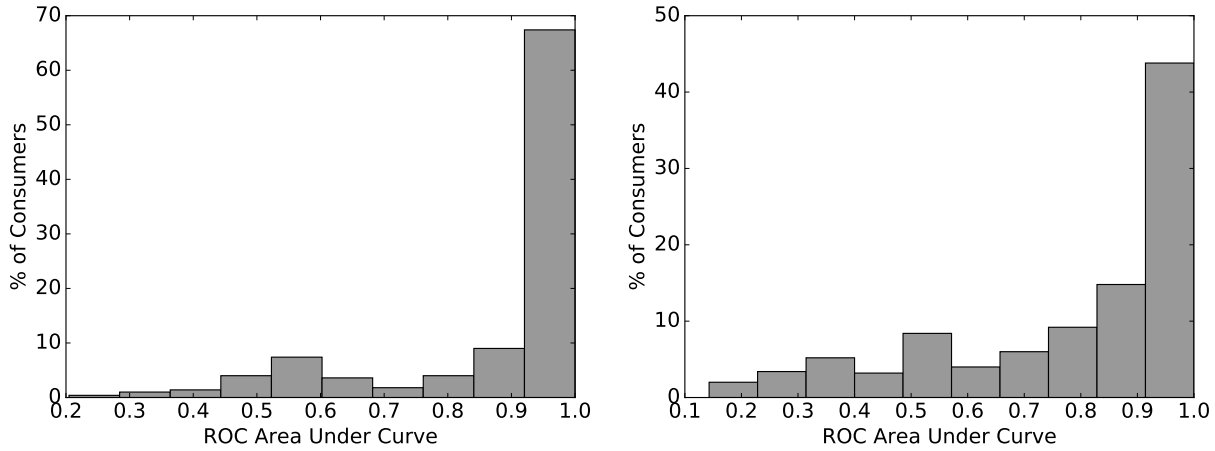
For the KLD detector, two parameters need to be chosen by a practitioner, and they determine the effectiveness of the detector. The first parameter is B, which is the number of bins that we would like to use to describe the nonparametric distribution of meter readings

in the training set. The second parameter to be chosen is the detection threshold on the distribution of KLD values. We had used percentiles to set the threshold, and had evaluated two choices (90^{th} and 95^{th} percentiles) in [95].

Figure B.1(a) illustrates the ROC curves for the KLD detector. The TPRs and FPRs were averaged over all consumers. The ROC curves are not monotonically increasing because B is based on a nonparametric distribution that is not smooth. As a result, there might be data points in the test set that exist in an empty bin, increasing the KLD metric because those points were not expected from the training set. How the bin size affects the TPRs and FPRs is more explicitly shown in Fig. B.1(b). It can be seen that a practitioner would do well to choose a bin size of 5 to achieve a good trade-off between TPRs and FPRs. With bin sizes of 4 and 6 there would be no major gains or losses over the choice of 5. In other words, the system is robust to small perturbations in bin count around 5.

Similarly, the operator would be well advised to pick a detection threshold at the 90^{th} percentile, so that the detector can operate at 87% TPR and 0.12% FPR. We believe that that may be an acceptable FPR, but the operator can pick a setting that achieves the best trade-off that they desire. The choice of percentile threshold is not immediately clear because the ROC curves for the different thresholds cross in Fig. B.1(a). One commonly used approach compares ROC curves based on the *area under the curves* (AUCs). A perfect detector has an AUC of 1. We do not see such perfect performance in the ROC curves shown in Fig. B.1 because the TPRs and FPRs were averaged over all consumers. To illustrate the performance of the detector for each consumer considered separately, we provide a histogram of AUCs in Fig. B.2(a). The AUCs were computed using the composite trapezoidal rule of integration.

Note that the FPR was dramatically decreased (in many cases by up to 20 percentage points) when the detector was "turned off" during the two weeks spanning Christmas and New Year's Day. That period produced the maximum number of false positives across all consumers, likely because consumption patterns were affected by holiday schedules.



(a) KLD area under ROC curve (90^{th} percentile) (b) PCA-DBSCAN area under ROC curve (3 dimensions)

Figure B.2: Area under the ROC curve (AUC) for KLD and PCA-DBSCAN detectors on the integrated ARIMA attack. The larger the area, the better the detection performance. For a large fraction of consumers, the detector had near-perfect performance (close to 1).

B.3 ROC for the PCA-DBSCAN Detector

We evaluate the PCA-DBSCAN detector against the integrated ARIMA attack in this section. The detector first projects the data into a lower-dimensional space. The dimensionality of that lower-dimensional space is a parameter that can be chosen by the operators. A smaller value will retain less of the variance in the original dataset. A larger value will include more noise from the dataset. Thus there is an optimal dimension for a given dataset, and we found that to be equal to 3 dimensions for the CER dataset, as illustrated in Fig. B.3. For each dimension, we generated the ROC curves by varying the ϵ parameter in DBSCAN, as described in Section 5.6.3. It is clear from Fig. B.3 that 3 dimensions work best because the corresponding curve lies entirely above the other curves. A histogram of AUCs for each consumer considered separately is plotted in Fig. B.2(b).

While it was clear from the ROC curves that the best KLD detector setting outperformed the best PCA-DBSCAN detector setting, we were surprised to find that both detectors had perfect performance on a large fraction of consumers, as seen in Fig. B.2. Upon investigating the consumers that performed poorly, we found that some of them had near-zero consumption throughout the 74-week period, while others had zero consumption during the period of

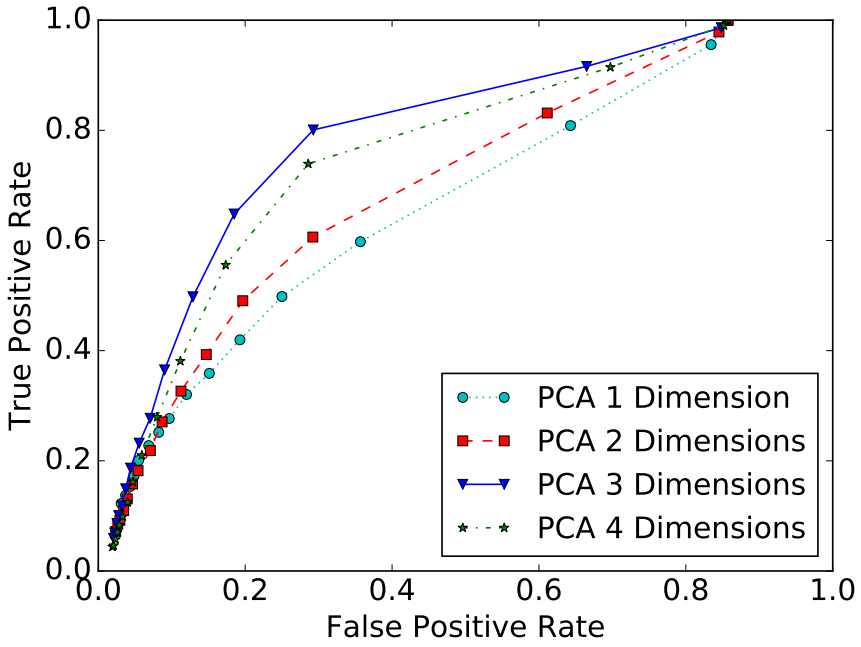


Figure B.3: ROC for PCA-DBSCAN detector on the integrated ARIMA attack.

the test set alone. Therefore, those consumers could not be distinguished from malicious consumers even to the naked eye, and they would need to be investigated by a utility. We spoke with a representative from the Pacific Gas & Electric company in California, and he told us that they use knowledge of move-in and move-out dates of residents and check that low consumption readings were seen after move-outs. If we had had access to those dates for the consumers in the CER dataset, we might have been able to further reduce the FPR.

APPENDIX C

CYBER ATTACKS ON REAL-TIME PRICING IN SMART GRIDS

In this appendix, we summarize the author's contribution to work on cyber-resilience in the context of electricity market systems. In particular, we consider real-time pricing (RTP) systems wherein consumers adapt their consumption in response to changing prices.

The signals sent by operators to consumers for demand response can take two broad forms. First, the signal can request a specified amount of demand reduction, as discussed in Chapter 3. For example, *Nest* allows a utility to remotely control HVAC loads in a house within specified constraints [159]. Second, the signal can represent a parameter that motivates the consumer to make a decision on whether to change their consumption. For example, if the electricity tariffs were to be suddenly raised dramatically, it may cause consumers to reduce their usage to avoid paying large electricity bills. In that sense, the DR signal influences a control decision made by the consumer as opposed to a control decision made by an operator.

In this appendix, we discuss the effects of cyber attacks on real-time electricity market prices, which are DR signals used by price-responsive consumers. We show that, in theory, such attacks could destabilize the electricity market system, and lead to consumption oscillations that could produce power outages that

The work in this appendix was originally published in [88] and later extended in [160]. This section presents the data-driven evaluation of cyber-attacks on RTP, which was the contribution of the author to [88] and [160]. The models for demand, generation, pricing, and the attacks were developed by Dr. Rui Tan, and they are briefly described to provide context for the author's work. Although the content of this appendix relates well with the cyber-resilience theme of this dissertation, the majority of the work was performed by Dr. Tan. Therefore, the work was not included in the main body of the dissertation.

C.1 Price-Responsive Demand

Unlike the work presented in the main content of the dissertation, which is empirical in nature, this work is theoretical. In this work, we assume that consumers have automated demand response (ADR) capabilities at their homes, which enable the automatic adaptation of consumption to the electricity price. Our consumption model is derived as follows.

Similar to (3.1) in Chapter 3, we split the load into a static baseline and a price-responsive component.

$$D(\lambda_k) = D_b(\lambda_k) + D_d(\lambda_k), \tag{C.1}$$

where b denotes the baseline load and d denotes the dynamic price-responsive load. For example, d could refer to an HVAC control system, like Nest, that was programmed to modify its consumption in response to price signals from a utility. Meanwhile, b may refer to appliances such as refrigerators that need to be constantly operating in order to keep food items from spoiling.

In our work we used the constant elasticity of own-price (CEO) model [161] to characterize the extent to which a consumer would modify their demand in response to changes in electricity price, λ. As per that model, the total price-responsive demand, D_d, is given as $D_d(\lambda_k) = \nu \cdot \lambda_k^{\epsilon}$, where ν and ϵ are positive and negative constants, respectively, and k is a time index for the time-varying price. The ϵ is referred to as the *price elasticity of demand*, which is typically within $(-1, 0)$ [162, 163].

C.2 Generation Model

In current electricity wholesale markets, the generation, denoted G, and price are determined through a bidding process [161], which is generally governed by the costs of generation and transmission. In a competitive bidding-based wholesale market, the price reflects the cost of electricity generation. We use a simplified linear function to reflect that cost of generation.

$$G(\lambda_k) = p\lambda_k + q, \tag{C.2}$$

where $p, q > 0$ are coefficients of the linear relationship that need to be determined by linear regression for each generation plant. The electricity pricing algorithm ensures that there is an agreement on how much a consumer is willing to pay and how much a generator is willing to receive as compensation for generation costs.

C.3 Pricing Algorithm

The RTP algorithm, proposed by Dr. Rui Tan in [88] assumes linear supply and demand models at a fixed operating point. It assumes that the closed loop formed by the generation and demand can be modeled as a linear time-invariant (LTI) system.

The advantage of the proposed RTP algorithm over a direct feedback approach proposed in [164] is that it ensures stability of the LTI system for any initial operating point, or any initial price λ_0. The price is updated at each time interval as follows.

$$\lambda_k = \lambda_{k-1} - \frac{2\eta}{\dot{G}(\lambda_{k-1}) - \dot{D}_d(\lambda_{k-1})} \cdot [G(\lambda_{k-1}) - D_d(\lambda_{k-1})], \tag{C.3}$$

where $\dot{G}$ operator denotes the first derivative of the generation with respect to the price, and similarly for the demand. η is an operator controlled setting that determines the convergence speed and resilience to attacks. Dr. Tan showed that $\eta = 0.5$ ensures maximum convergence speed when the baseline load is constant. Lower values result in lower convergence speeds, but greater resilience to attacks. We refer the interested reader to [88] for details on the derivation of the algorithm.

C.4 Attack Model

In this chapter we consider attacks that compromise the integrity of pricing signals, with the intention of destabilizing the grid and causing power outages. The assumption is that the electricity price is communicated to consumers through computer communication networks, and consumers have automated DR mechanisms that cause their consumption to adapt to every change in price, in accordance with the CEO model.

There are usually a large number of consumers in any electric grid. In order to destabilize the grid, an attacker would need to compromise the price that is seen by a significant fraction of those consumers. Let C denote the set of all consumers and $C' \subseteq C$ denote the set of consumers whose price signals are compromised. Let $D_{d,j}$ denote the price-responsive demand for consumer j, where $j \in C$. We define

$$\rho = \frac{\sum_{j \in C'} D_{d,j}}{\sum_{j \in C} D_{d,j}}, \tag{C.4}$$

which characterizes the fraction of demand that has been compromised by the integrity attack on electricity price.

In this appendix, we consider integrity attacks wherein the malicious modifications follow certain rules and can be achieved by an attacker with lower capability and resource requirements. We assume that the compromised price, denoted by λ'_k, is seen by all consumers in the set $\{j | j \in C'\}$. In other words, there is only one compromised price; the consumers who are not compromised see the true price λ_k. An attack can be characterized by the parameters for the rule, which is denoted by $\mathcal{A}$. We consider two kinds of integrity attacks:

Scaling attack $\mathcal{A} = (\rho, \gamma)$: The compromised price is a scaled version of the true price, i.e., $\lambda'_k = \gamma \lambda_k$, $\gamma \in \mathbb{R}^+$.

Delay attack $\mathcal{A} = (\rho, \tau)$: The compromised price is an old price, i.e., $\lambda'_k = \lambda_{k-\tau}$, $\tau \in \mathbb{Z}^+$.

These two attacks can be launched in various ways. For example, the price values or timestamps in data packets sent to the smart meters can be maliciously modified during transmissions in vulnerable communication networks. The delay attack can also be launched by modifying the smart meters' internal clocks. Smart meters typically assign a memory buffer to store received prices. If a smart meter's clock has a lag, it will store newly received prices in the buffer and apply an old price for the present. Furthermore, attacks on the clocks can be realized by compromising vulnerable time synchronization protocols or the time servers that provide timing information to the smart meters. A few smart meter products [165] synchronize their clocks via a built-in GPS receiver, which is vulnerable and subject to remote attacks that are effective across large geographic areas [166].

In this section, we assume that at most one kind of attack is in effect. That simplification

allows us to better understand the impact of each attack on the RTP system, which is the basis for understanding more complex scenarios such as combinations of attack types.

C.5 Attack Simulations

We use GridLAB-D [167], an electric power distribution system simulator, to evaluate the impact of the attacks described in Section C.4. GridLAB-D models steady-state power flows in the distribution grid, while taking into account line impedances, transformer losses and thermal capacities of electric conductors. Therefore, a simulation using GridLAB-D relaxes many of the simplifying assumptions made in Dr. Tan's control-theoretic analysis of the attacks in [88]. Furthermore, GridLAB-D records emergency events that occur when the thermal ratings of distribution lines and transformers are exceeded. Those events could cause sustained disruption of power delivery to consumers. Those events are of particular interest to us because they help us understand the physical consequences of the attacks.

C.5.1 Simulation Methodology and Settings

We use a distribution feeder specification [168], which covers a moderately populated urban area and comprises 1405 houses, 2134 buses, 3314 triplex buses, 1944 transformers, 1543 overhead lines, 335 underground lines, and 1631 triplex lines. For this small-scale distribution feeder, locational prices are usually not applicable and hence all the houses are subject to the same price. As GridLAB-D did not provide real-time pricing market features, we extended the open-source simulator's capabilities by developing new modules for it. The *demand module* implemented the CEO model for each house, the *ISO module* implemented the price stabilization algorithm in Eq. (C.3), and the *adversary module* implemented the two attack strategies. Each module could be configured with appropriate parameters using GridLAB-D's simulation configuration language. We automatically generated new configurations and ran the simulator to evaluate the impact of a wide spectrum of attack parameters.

We measured the instantaneous power of the entire feeder at the root node. Its peak value over the previous pricing period is used as $D(\lambda_{k-1})$ in Eq. (C.3). As we focus on evaluating

the physical consequences of attacks, we do not simulate the logistics of the attacks and assume that the adversary can gain access to the meters of his choosing. Specifically, if a house is not subject to attacks, it directly reads the real-time price from the ISO module; otherwise, it reads the price from the adversary module. All the attacks are launched after the real-time pricing algorithm has converged.

We adopt the CEO demand model for each single house, where the parameters are drawn from normal distributions: $\nu \sim \mathcal{N}(7, 3.5^2)$ (unit: kW) and $\epsilon \sim \mathcal{N}(-0.8, 0.1^2)$. Under this setting, if the price is within $[10, 20]$, the per-house price-responsive demand is within $[0.65, 1.1]$ kW. To improve the realism of the simulations, we use the half-hourly total demand of NSW, Australia, provided by AEMO [169] as the baseline load. The data obtained was from the period March 1st to 22nd, 2013. The baseline load of a single house is set to be a scaled version of the real data. The resultant range of the per-house baseline load is $[0.276, 0.488]$ kW. Hence, when the price is within $[10, 20]$, the demand of a household is within $[0.9, 1.6]$ kW, which is consistent with the average demand of a household in reality [170]. In our simulations, the price is updated every half hour, to be consistent with the demand data obtained from [169]. In each pricing period, the simulated demand remains constant. For the generation model in Eq. (C.2), we set $p = 0.044$ and $q = 1.287$ to be commensurate with the total demand from the dataset. Other default settings include: $T = 0.5$, $\lambda_{\min} = 1$, $\lambda_{\max} = 200$, and $\eta = 0.5$.

C.5.2 Simulation Results

Price stabilization

The first set of simulations evaluates the effectiveness of the direct feedback approach [164] and our control-theoretic price stabilization algorithm in Eq. (C.3). In the simulations, the direct feedback approach is unstable, where the price oscillates between $\lambda_{\min}$ and $\lambda_{\max}$. The total demand reaches 10 MW a few hours after the start of the simulation, and GridLAB-D reports that four distribution lines are overloaded. Figure C.1 plots the price and resultant demands under the price stabilization approach in Eq. (C.3). We can see that the price

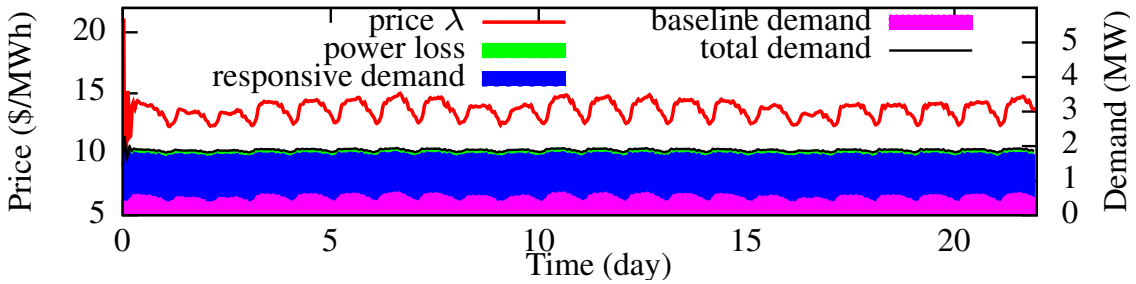


Figure C.1: Price stabilization in the absence of attack.

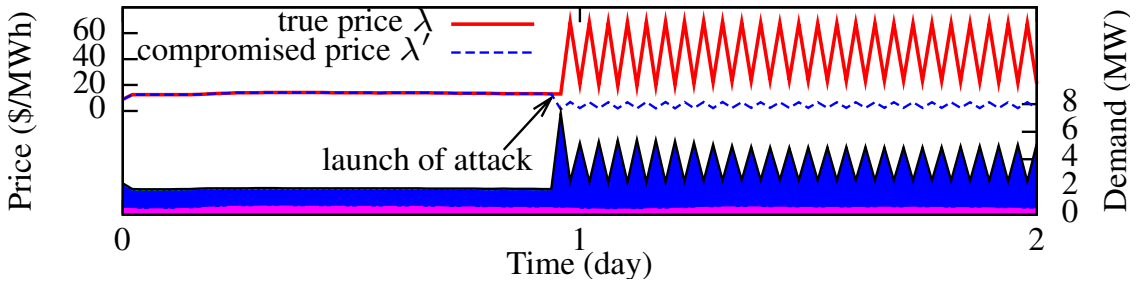
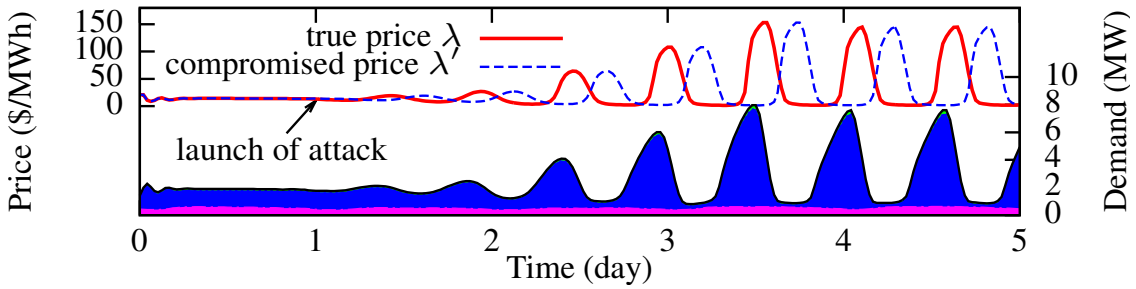


Figure C.2: Scaling attack ($\rho = 65\%$, $\gamma = 0.1$).

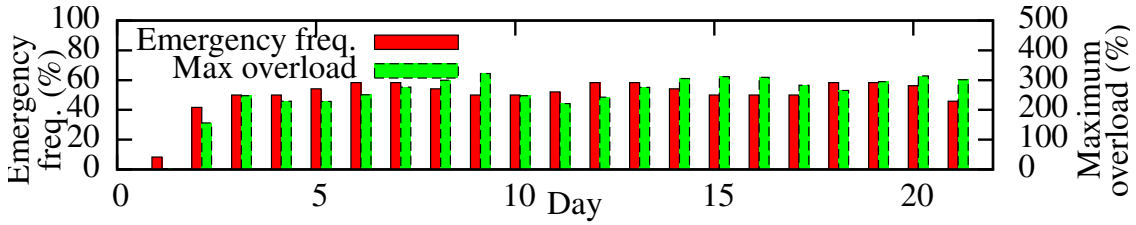
fluctuates slightly for a few hours after the start of the simulation, due to an inappropriate initial price. After the system converged, it adapted to the time-varying baseline load. The generation scheduling error is close to zero, which means that the clearing price is achieved.

Scaling attack

Figure C.2 plots the true and compromised prices, as well as the breakdown of demand under the scaling attack. We can see that the price as well as the demand fluctuates severely. GridLAB-D reports excessive distribution line overload events after the launch of the attack. We also extensively evaluate the impact of the scaling attack with different settings of ρ and γ. We use the standard deviation of the generation scheduling error after the launch of the attack, denoted by $\sigma(e)$, as the system volatility metric. A near-zero $\sigma(e)$ means convergence, while a considerably large $\sigma(e)$ means oscillation or divergence. Figure C.5a plots $\sigma(e)$ versus γ under various settings of ρ. We can see that the system volatility increases with ρ and decreases with γ.



(a) Prices and demands



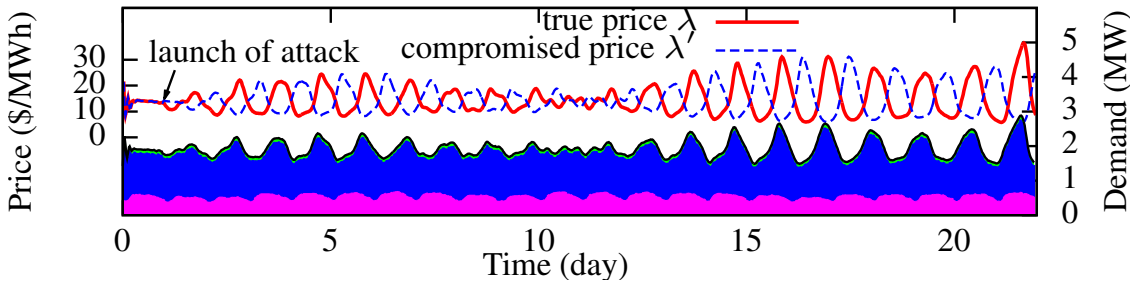
(b) Emergencies over time

Figure C.3: Impact of delay attack ($\rho = 100\%$, $\tau = 9$).

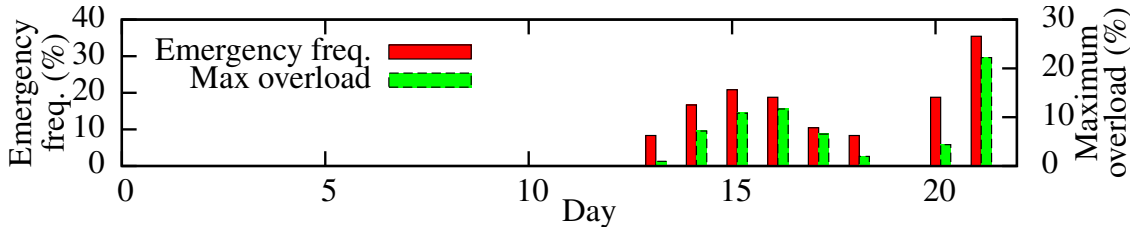
Delay attack

Figure C.3(a) and Fig. C.4(a) show the evolution of price and the breakdown of demand under the delay attacks with different parameters. We also investigate the emergency events reported by GridLAB-D. The *overload* of a distribution line or a transformer is defined as the percentage of the exceeded current/power with respect to the rated value. Figure C.3(b) and Fig. C.4(b) plot the *emergency frequency* and maximum overload in each day. The emergency frequency is defined as the ratio of the number of pricing periods with reported emergency events to the number of pricing periods per day (i.e., 48). In Fig. C.3, a small generation scheduling error caused by the time-varying baseline load will be amplified iteratively along the control loops, after the launch of the attack. The overload can be up to 350%. In practice, such a high overload will cause circuit breakers to open and lead to regional blackouts. In Fig. C.4, the system appears to diverge and then converge again without causing any emergencies. However, it diverges again from the 12th day due to the changing baseline load, causing excessive emergencies. This behavior illustrates the stealthiness of the delay attack that causes marginal system stability.

Finally, we evaluated the impact of the delay attack with different settings of ρ and τ. The results are shown in Fig. C.5b. We can clearly see that when $\rho < 0.5$, the system remains



(a) Prices and demands



(b) Emergencies over time

Figure C.4: Impact of delay attack ($\rho = 65\%$, $\tau = 24$).

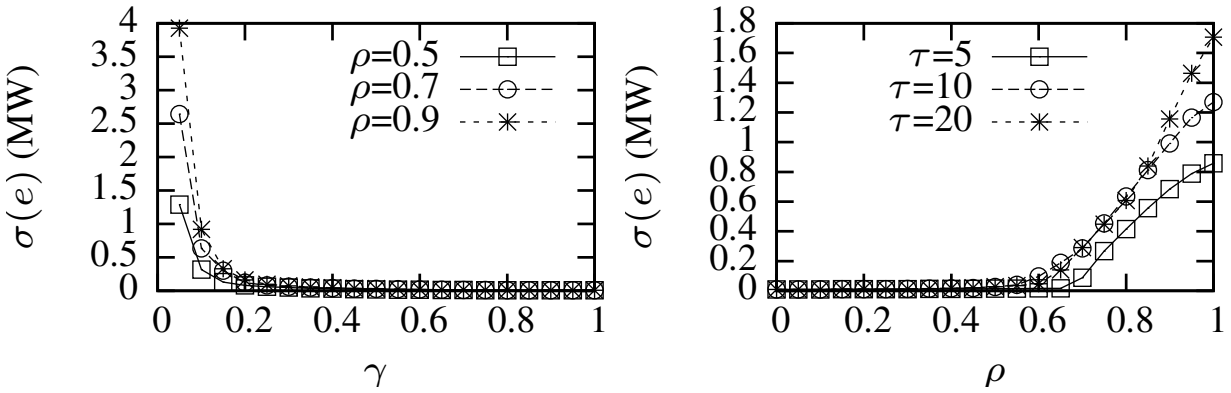
stable, which is consistent with the theoretical result obtained by Dr. Tan in [88].

C.6 Related Work

Related work does not consider the response of consumers to prices. In [147], Liu et al. examine the conditions for bypassing false data injection detectors. The authors of [171–175] showed that the false data injection attacks can lead to increased system operation costs due to inordinate generation dispatch [171] or energy routing [172], as well as economic losses due to misconduct of electricity markets [173–175]. In particular, the studies in [173–175] focus on false data injection attacks on real-time wholesale markets.

C.7 Conclusion

We evaluated integrity attacks on DR pricing signals using the GridLAB-D simulator. Although the simulation relaxes many of the assumptions that Dr. Tan made in his theoretical analysis of the attacks on RTP, it supports his findings that the attacker would need to compromise at least half of the demand in the grid in order to create price fluctuations and



(a) Scaling attack

(b) Delay attack

Figure C.5: System volatility under attacks.

instability.

REFERENCES

[1] J. Weeks, *U.S. Electrical Grid Undergoes Massive Transition to Connect to Renewables*, Scientific American, April 2010, https://www.scientificamerican.com/article/what-is-the-smart-grid/.

[2] W. H. Sanders, "Construction and solution of performability models based on stochastic activity networks," Ph.D. Thesis, University of Michigan, 1988. [Online]. Available: http://www.perform.illinois.edu/Papers/USAN_papers/88S02.pdf

[3] J. Miller, C. Folgar, and J. McCuan, *The fastest-growing area of machine learning science*, Quid, March 2017, https://quid.com/feed/the-fastest-growing-area-of-machine-learning-science.

[4] *Large Scale Visual Recognition Challenge (ILSVRC)*, ImageNet, http://www.image-net.org/challenges/LSVRC/. Accessed on October 2018.

[5] Q. Ye, J. Lu, and M. Zhu, "Wind curtailment in China and lessons from the United States," March 2018, Brookings. [Online]. Available: https://www.brookings.edu/research/wind-curtailment-in-china-and-lessons-from-the-united-states/

[6] M. Joos and I. Staffell, "Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany," *Renewable and Sustainable Energy Reviews*, vol. 86, pp. 45 – 65, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032118300091

[7] V. Singh, "Power theft continues to hit Indian economy," July 2018, Media India Group. [Online]. Available: https://mediaindia.eu/business-politics/power-theft-continues-to-hit-indian-economy/

[8] BC Hydro, "Smart meters help reduce electricity theft, increase safety," May 2011. [Online]. Available: https://www.bchydro.com/news/conservation/2011/smart_meters_energy_theft.html

[9] M. Ward, "Smart meters can be hacked to cut power bills," October 2014. [Online]. Available: http://www.bbc.com/news/technology-29643276

[10] Cyber Intelligence Section, "Smart grid electric meters altered to steal electricity," May 2010, Federal Bureau of Investigation. [Online]. Available: http://krebsonsecurity.com/wp-content/uploads/2012/04/FBI-SmartMeterHack-285x305.png

[11] K. Zetter, *Simulated Cyberattack Shows Hackers Blasting Away at the Power Grid*, Wired, September 2007, available https://www.wired.com/2007/09/simulated-cyber/. Accessed October 2018.

[12] K. Zetter, *An Unprecedented Look at Stuxnet, the World's First Digital Weapon*, Wired, March 2014, available https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/. Accessed December 2017.

[13] A. Greenberg, *'Crash Override': The Malware That Took Down a Power Grid*, Wired, June 2017, available https://www.wired.com/story/crash-override-malware/. Accessed December 2017.

[14] M. Kenderdine, *US power grid needs defense against looming cyber attacks*, The Hill, March 2018, https://thehill.com/opinion/energy-environment/379980-us-power-grid-needs-defense-against-looming-cyber-attacks.

[15] L. Yuanyuan, "Wind power curtailment in China expected to increase in second half of 2016," August 2016, RenewableEnergyWorld.com. [Online]. Available: https://www.renewableenergyworld.com/articles/2016/08/wind-power-curtailment-in-china-expected-to-increase-in-second-half-of-2016.html

[16] *Renewables 2016 Global Status Report*, Renewable Energy Policy Network for the 21st Century (REN21), 2016, http://www.ren21.net/wp-content/uploads/2016/06/GSR_2016_Full_Report.pdf.

[17] R. Wiser and M. Bolinger, *2015 Wind Technologies Market Report*, US Department of Energy, August 2016, https://energy.gov/sites/prod/files/2016/08/f33/2015-Wind-Technologies-Market-Report-08162016.pdf.

[18] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Trans. on Energy Conversion*, vol. 21, no. 1, pp. 273–284, March 2006.

[19] A. D. Piazza, M. D. Piazza, and G. Vitale, "Estimation and forecast of wind power generation by FTDNN and NARX-net based models for energy management purpose in smart grids," *Renewable Energies and Power Quality Journal (RE&PQJ)*, vol. 1, no. 12, April 2014.

[20] S. Kurandwad, C. Subramanian, V. R. P, A. Vasan, V. Sarangan, V. Chellaboina, and A. Sivasubramaniam, "Windy with a chance of profit: Bid strategy and analysis for wind integration," in *Proceedings of the 5th International Conference on Future Energy Systems*, ser. e-Energy '14. New York, NY, USA: ACM, 2014, pp. 39–49.

[21] B. Narayanaswamy, V. K. Garg, and T. S. Jayram, "Online optimization for the smart (micro) grid," in *Proceedings of the 3rd International Conference on Future Energy Systems*, ser. e-Energy '12. New York, NY, USA: ACM, 2012, pp. 19:1–19:10.

[22] M. H. Hajiesmaili, C.-K. Chau, M. Chen, and L. Huang, "Online microgrid energy generation scheduling revisited: The benefits of randomization and interval prediction," in *Proceedings of the Seventh International Conference on Future Energy Systems*, ser. e-Energy '16. New York, NY, USA: ACM, 2016, pp. 1:1–1:11.

[23] J. Taneja, K. Lutz, and D. Culler, "Flexible loads in future energy networks," in *Proceedings of the Fourth International Conference on Future Energy Systems*, ser. e-Energy '13. New York, NY, USA: ACM, 2013, pp. 285–286.

[24] L. Gan, A. Wierman, U. Topcu, N. Chen, and S. H. Low, "Real-time deferrable load control: Handling the uncertainties of renewable generation," in *Proceedings of the Fourth International Conference on Future Energy Systems*, ser. e-Energy '13. New York, NY, USA: ACM, 2013, pp. 113–124.

[25] *Weather Research And Forecasting Model*, National Center for Atmospheric Research (NCAR), https://www.mmm.ucar.edu/weather-research-and-forecasting-model Accessed January 2018.

[26] A. Kusiak, H. Zheng, and Z. Song, "Short-term prediction of wind farm power: A data mining approach," *IEEE Transactions on Energy Conversion*, vol. 24, no. 1, pp. 125–136, March 2009.

[27] R. Blonbou, "Very short-term wind power forecasting with neural networks and adaptive bayesian learning," *Elsevier Renewable Energy*, vol. 36, pp. 1118–1124, August 2010.

[28] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[29] *Getting started with the Keras Sequential model*, Keras, available https://keras.io/getting-started/sequential-model-guide/. Accessed December 2017.

[30] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium (NAPS), 2010*. IEEE, 2010, pp. 1–8.

[31] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Applied Energy*, vol. 98, pp. 574 – 583, 2012.

[32] S. Shokrzadeh, M. J. Jozani, and E. Bibeau, "Wind turbine power curve modeling using advanced parametric and nonparametric methods," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 4, pp. 1262–1269, Oct 2014.

[33] V. B. Krishna, C. A. Gunter, and W. H. Sanders, "Evaluating detectors on optimal attack vectors that enable electricity theft and der fraud," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 4, pp. 790–805, Aug 2018.

[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1," D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Eds. Cambridge, MA, USA: MIT Press, 1986, ch. Learning Internal Representations by Error Propagation, pp. 318–362. [Online]. Available: http://dl.acm.org/citation.cfm?id=104279.104293

[35] *Winter storm hits frozen East Coast*, National Weather Service, May 2015. [Online]. Available: http://www.weather.gov/media/ilx/Stormdata/may2015.pdf

[36] C. Dolce, "Storm brings strong winds, heavy rain, power outages and some snow to midwest, northeast," October 2015. [Online]. Available: https://weather.com/travel/commuter-conditions/news/eastern-storm-rain-wind-snow

[37] C. AP, "Winter storm hits frozen east coast," February 2016. [Online]. Available: https://www.cbsnews.com/news/winter-storm-hits-frozen-east-coast-possible-tornadoes-in-the-south/

[38] A. Lahouar and J. B. H. Slama, "Hour-ahead wind power forecast based on random forests," *Renewable Energy*, vol. 109, pp. 529 – 541, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148117302550

[39] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1 – 8, 2012.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[42] T. G. Barbounis and J. B. Theocharis, "Locally recurrent neural networks for wind speed prediction using spatial correlation," *Elsevier Information Sciences*, vol. 177, pp. 5775–5797, May 2007.

[43] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[44] *LinearSVR*, Scikit Learn, http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVR.html Accessed January 2018.

[45] W. Ningbo, L. Liang, L. Guangtu, W. Dingmei, and L. Qingquan, "Ultrashort-term prediction method for wind electricity power of arma model with anemometer network real-time correction," Aug. 13 2014, CN Patent App. CN 201,410,187,023.

[46] R. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for r," *Journal of Statistical Software*, vol. 27, no. 1, pp. 1–22, 2008.

[47] V. Badrinath Krishna, R. K. Iyer, and W. H. Sanders, "ARIMA-Based Modeling and Validation of Consumption Readings in Power Grids," in *Proceedings of Critical Information Infrastructure Security (CRITIS)*. Springer Verlag, 2015.

[48] P. Pinson and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models," *Journal of Forecasting*, vol. 31, no. 4, pp. 281–313, 2012.

[49] M. Negnevitsky, P. Mandal, and A. K. Srivastava, "Machine learning applications for load, price and wind power prediction in power systems," in *2009 15th International Conference on Intelligent System Applications to Power Systems*, Nov 2009, pp. 1–6.

[50] E. Cadenas, W. Rivera, R. Campos-Amezcua, and R. Cadenas, "Wind speed forecasting using the NARX model, case: La Mata, Oaxaca, México," *Neural Computing and Applications*, vol. 27, no. 8, pp. 2417–2428, Nov 2016.

[51] Z. Men, E. Yee, F. Lien, Z. Yang, and Y. Liu, "Ensemble nonlinear autoregressive exogenous artificial neural networks for short-term wind speed and power forecasting," *International Scholarly Research Notices*, vol. 2014, no. 972580, 2014.

[52] J. Kleissl, *Solar Energy Forecasting and Resource Assessment*. Elsevier Press, 2013.

[53] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. IEEE, 2011, pp. 528–533.

[54] S. Iyengar, N. Sharma, D. Irwin, P. Shenoy, and K. Ramamritham, "Solarcast: a cloud-based black box solar predictor for smart homes," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014, pp. 40–49.

[55] D. Chen and D. Irwin, "Sundance: Black-box behind-the-meter solar disaggregation," in *Proceedings of the Eighth International Conference on Future Energy Systems*. ACM, 2017, pp. 45–55.

[56] S. Mathew, J. Hazra, S. A. Husain, C. Basu, L. C. DeSilva, D. Seetharam, N. Y. Voo, S. Kalyanaraman, and Z. Sulaiman, "An advanced model for the short-term forecast of wind energy," in *MODSIM 2011-19th International Congress on Modelling and Simulation-Sustaining Our Future: Understanding and Living with Uncertainty*. MSSANZ, 2011, pp. 1745–1752.

[57] V. B. Krishna, W. S. Wadman, and Y. Kim, "NowCasting: Accurate and precise short-term wind power prediction using hyperlocal wind forecasts," in *Proceedings of the Ninth International Conference on Future Energy Systems*, ser. e-Energy '18. New York, NY, USA: ACM, 2018. [Online]. Available: http://doi.acm.org/10.1145/3208903.3208919 pp. 63–74.

[58] D. Jung, V. Badrinath Krishna, K. Ngo Quang Minh, H. H. Nguyen, and D. K. Y. Yau, "EnergyTrack: Sensor-Driven Energy Use Analysis System," in *ACM BuildSys*, 2013.

[59] D. Jung, V. Badrinath Krishna, W. G. Temple, and D. K. Yau, "Data-driven evaluation of building demand response capacity," in *Proceedings of IEEE SmartGridComm'14*. IEEE, 2014, pp. 547–553.

[60] Efficiency Valuation Organization, "International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings," *ISO, Geneva, Switzerland*, vol. 1, 2012.

[61] A. Marszal, P. Heiselberg, J. Bourrelle, E. Musall, K. Voss, I. Sartori, and A. Napolitano, "Zero Energy Building - A review of definitions and calculation methodologies," *Energy and Buildings*, vol. 43, no. 4, pp. 971 – 979, 2011.

[62] "EnergyPlus," http://apps1.eere.energy.gov/buildings/energyplus/energyplus_about.cfm.

[63] A. Boyano, P. Hernandez, and O. Wolf, "Energy demands and potential savings in european office buildings: Case studies based on energyplus simulations," *Energy and Buildings*, vol. 65, no. 0, pp. 19 – 28, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378778813003277

[64] ISO 7730:2005, "Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria," 2005.

[65] M. Stewart, "On least squares estimation when the dependent variable is grouped," *Review of Economic Studies*, vol. 50, no. 4, pp. 737–53, 1983.

[66] "Guide to providing interruptible load in singapore's wholesale electricity market," https://www.emcsg.com/f146,16653/Guide_to_providing_IL_website_21082012.pdf.

[67] "Sample Demand Response Audit," http://www.nationalgridus.com/non_html/shared_demand_response_sample_audit.pdf.

[68] "Demand Response Measurement and Verification," https://www.smartgrid.gov/sites/default/files/pdfs/demand_response.pdf.

[69] J. Mathieu, P. Price, S. Kiliccote, and M. Piette, "Quantifying Changes in Building Electricity Use, With Application to Demand Response," *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp. 507–518, 2011.

[70] N. Motegi, M. A. Piette, D. S. Watson, S. Kiliccote, and P. Xu, "Introduction to Commercial Building Control Strategies and Techniques for Demand Response," Lawrence Berkeley National Laboratory, Tech. Rep. 59975, 2007.

[71] "Demand Response Quick Assessment Tool," http://drrc.lbl.gov/drqat.

[72] D. B. Arnold, M. D. Sankur, and D. M. Auslander, "Optimal Control of Office Plug-Loads for Commercial Building Demand Response," in *ASME 2013 Dynamic Systems and Control Conference*, 2013.

[73] S. Dawson-Haggerty, S. Lanzisera, J. Taneja, R. Brown, and D. Culler, "@ scale: Insights from a large, long-lived appliance energy WSN," in *IPSN*, 2012.

[74] N. Murthy, "Energy-Agile Laptops: Demand Response of Mobile Plug Loads Using Sensor/Actuator Networks," in *IEEE SmartGridComm*, 2012.

[75] V. L. Erickson and A. E. Cerpa, "Occupancy Based Demand Response HVAC Control Strategy," in *ACM BuildSys*, 2010.

[76] S. Raziei and H. Mohsenian-Rad, "Optimal Demand Response Capacity of Automatic Lighting Control," in *IEEE ISGT*, 2013.

[77] Y. Agarwal, B. Balaji, S. Dutta, R. Gupta, and T. Weng, "Duty-Cycling Buildings Aggressively: The Next Frontier in HVAC Control," in *IPSN*, 2011.

[78] R. Jiang, R. Lu, L. Wang, J. Luo, S. Changxiang, and S. Xuemin, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Science And Technology*, vol. 19, no. 2, pp. 105–120, April 2014.

[79] R. Katakey and R. K. Singh, "India fights to keep the lights on," June 2014, Bloomberg Businessweek. [Online]. Available: http://www.businessweek.com/printer/articles/205322-india-fights-to-keep-the-lights-on

[80] BC Hydro, "Smart metering program," 2014, BC Hydro. [Online]. Available: https://www.bchydro.com/energy-in-bc/projects/smart_metering_infrastructure_program.html

[81] Coalition to Stop Mart Meters in BC, "Theft of power through hacking of smart meters," October 2014. [Online]. Available: http://www.stopsmartmetersbc.com/2014-10-23-theft-of-power-through-hacking-of-smart-meters/

[82] J. D. Glover, M. Sarma, and T. Overbye, *Power System Analysis & Design, SI Version*. Cengage Learning, 2011.

[83] E. de Buda, "System for accurately detecting electricity theft," U.S. Patent 12/351,978, 2010.

[84] D. N. Nikovski, Z. Wang, A. Esenther, H. Sun, K. Sugiura, T. Muso, and K. Tsuru, "Smart meter data analysis for power theft detection," in *Proceedings MLDM'13*. Springer-Verlag, 2013, pp. 379–389.

[85] Edison Electric Institute, "Smart meters and smart meter systems: A metering industry perspective," p. 35, March 2011. [Online]. Available: http://www.eei.org/issuesandpolicy/grid-enhancements/documents/smartmeters.pdf

[86] D. Mashima and A. A. Cárdenas, "Evaluating electricity theft detectors in smart grid networks," in *Proceedings of RAID'12*, vol. 7462. Springer, 2012, pp. 210–229.

[87] OASIS Consortium, "Energy Market Information Exchange (EMIX) Version 1.0," January 2012. [Online]. Available: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=emix

[88] R. Tan, V. Badrinath Krishna, D. K. Yau, and Z. Kalbarczyk, "Impact of integrity attacks on real-time pricing in smart grids," in *Proceedings of ACM CCS'13*. New York, NY, USA: ACM, 2013, pp. 439–450.

[89] V. Badrinath Krishna, G. A. Weaver, and W. H. Sanders, "PCA-Based Method for Detecting Integrity Attacks on Advanced Metering Infrastructure," in *Proceedings of Quantitative Evaluation of Systems (QEST)*. Springer Verlag, 2015, pp. 70–85.

[90] S. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," in *Proceedings of IEEE PSCE'11*, March 2011, pp. 1–8.

[91] S. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi, "Measures and setbacks for controlling electricity theft," in *Proceedings of North American Power Symposium (NAPS)*, Sept 2010, pp. 1–8.

[92] P. Kelly-Detwiler, "Electricity theft: A bigger issue than you think," April 2014, Forbes. [Online]. Available: http://www.forbes.com/sites/peterdetwiler/2013/04/23/electricity-theft-a-bigger-issue-than-you-think/

[93] S. McLaughlin, D. Podkuiko, S. Miadzvezhanka, A. Delozier, and P. McDaniel, "Multi-vendor penetration testing in the advanced metering infrastructure," in *Proceedings of ACSAC'10*. New York, NY, USA: ACM, 2010, pp. 107–116.

[94] S. McWilliams, "Tamper protection for an automatic remote meter reading unit," U.S. Patent 4,357,601, 1982, Bell Telephone Laboratories.

[95] V. Badrinath Krishna, K. Lee, G. A. Weaver, R. K. Iyer, and W. H. Sanders, "F-deta: A framework for detecting electricity theft attacks in smart grids," in *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, June 2016, pp. 407–418.

[96] ComEd, "Safeguarding data through smarter technology," 2014, Commonwealth Edison Company. [Online]. Available: https://www.comed.com/documents/technology/grid_mod_fact_sheet_security_2014_r2.pdf

[97] P. Jovanovic and S. Neves, "Practical cryptanalysis of the open smart grid protocol," in *Proceedings of Fast Software Encryption*. Springer Berlin Heidelberg, 2015, vol. 9054, pp. 297–316.

[98] S. McLaughlin, B. Holbert, S. Zonouz, and R. Berthier, "AMIDS: A multi-sensor energy theft detection framework for advanced metering infrastructures," in *Proceedings of SmartGridComm*, 2012, pp. 354–359.

[99] *Worlds first open source framework allows individuals to test for vulnerabilities in Smart Meters*, SecureState Consulting, June 2012, https://www.securestate.com/blog/2012/06/27/termineter-framework-open-source-smart-meter-hacking-tool.

[100] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting : the forecast package for R Automatic time series forecasting : the forecast package for R," *Journal Of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.

[101] "Forecasting with ARIMA Models." [Online]. Available: https://onlinecourses.science.psu.edu/stat510/node/66

[102] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Proceedings of KDD'96*, vol. 96, 1996, pp. 226–231.

[103] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering Points to Identify The Clustering Structure," *ACM SIGMOD Record*, vol. 28, no. 2, June 1999.

[104] P. J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, 1993.

[105] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in *Critical Information Infrastructures Security*, E. Rome and R. Bloomfield, Eds. Springer Berlin Heidelberg, 2010, vol. 6027, pp. 176–187.

[106] M. Afgani, S. Sinanovic, and H. Haas, "Anomaly detection using the Kullback-Leibler divergence metric," in *proceedings of ISABEL'08*, Oct 2008, pp. 1–5.

[107] J. Harmouche, C. Delpha, and D. Diallo, "Faults diagnosis and detection using principal component analysis and kullback-leibler divergence," in *In proceedings of IEEE IECON'12*, Oct 2012, pp. 3907–3912.

[108] R. Fu, D. Chung, T. Lowder, D. Feldman, K. Ardani, and R. Margolis, *U.S. Solar Photovoltaic System Cost Benchmark: Q1 2016*, National Renewable Energy Laboratory (NREL), September 2016, http://www.nrel.gov/docs/fy16osti/66532.pdf.

[109] A. Lee, *U.S. wind generating capacity surpasses hydro capacity at the end of 2016*, U.S. Energy Information Administration, accessed May 2017. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=30212

[110] *Rising solar generation in California coincides with negative wholesale electricity prices*, U.S. Energy Information Administration, April 2017, accessed May 2017. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=30692

[111] *Renewable Energy Technologies: Cost Analysis Series (Wind Power)*, International Renewable Energy Agency, June 2012, accessed May 2017. [Online]. Available: https://www.irena.org/DocumentDownloads/Publications/RE_Technologies_Cost_Analysis-WIND_POWER.pdf

[112] P. Bonami, L. T. Biegler, A. R. Conn, G. Cornuejols, I. E. Grossmann, C. D. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, and A. Wachter, "An algorithmic framework for convex mixed integer nonlinear programs," *Discrete Optimization*, vol. 5, no. 2, pp. 186 – 204, 2008, in Memory of George B. Dantzig. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1572528607000448

[113] P. Bonami, M. Kilinç, and J. Linderoth, "Algorithms and software for convex mixed integer nonlinear programs," in *Mixed Integer Nonlinear Programming*, J. Lee and S. Leyffer, Eds. New York, NY: Springer New York, 2012, pp. 1–39.

[114] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "SCS: Splitting conic solver, version 2.0.2," Nov. 2017. [Online]. Available: https://github.com/cvxgrp/scs

[115] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Machine Learning Res.*, vol. 17, no. 83, pp. 1–5, 2016.

[116] *Solar Home Electricity Data*, Ausgrid, July 2010, http://www.ausgrid.com.au/Common/About-us/Corporate-information/Data-to-share/Solar-home-electricity-data.aspx.

[117] National Renewable Energy Laboratory, *Solar Power Data*, December 2016, http://www.nrel.gov/grid/solar-power-data.html.

[118] *Deep Thunder*, IBM, August 2003. [Online]. Available: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/

[119] *The Wind Turbine Database*, WindPower Program, http://www.wind-power-program.com/download.htm\#database. Accessed May 2017.

[120] *PVWatts Calculator*, National Renewable Energy Laboratory (NREL), accessed April 2017. [Online]. Available: http://pvwatts.nrel.gov/pvwatts.php

[121] *Setting a fair and reasonable value for electricity generated by small-scale solar PV units in NSW*, Independent Pricing and Regulatory Tribunal, March 2012, https://www.ipart.nsw.gov.au/files/sharedassets/website/trimholdingbay/final_report_-_solar_feed-in_tariffs_-_march_2012.pdf.

[122] K. Shallenberger, *CAISO: Wholesale power prices dropped 9% in 2016 to $34/MWh average*, UtilityDive, May 2017, accessed October 2017. [Online]. Available: http://www.utilitydive.com/news/caiso-wholesale-power-prices-dropped-9-in-2016-to-34mwh-average/442626/

[123] *Epex Spot power exchange*, RTE, accessed October 2017. [Online]. Available: https://clients.rte-france.com/lang/an/visiteurs/vie/marche_electricite.jsp

[124] Solar Choice Pty Ltd, *1.5kW solar PV systems: Pricing, outputs and payback*, August 2016, http://www.solarchoice.net.au/blog/1-5kw-solar-pv-systems-price-output-payback.

[125] Z. Sheftalovich, *Which solar power system should you get?*, January 2015, https://www.choice.com.au/home-improvement/energy-saving/solar/articles/solar-power-survey-results.

[126] *How much do wind turbines cost?*, Windustry, http://www.windustry.org/how_much_do_wind_turbines_cost. Accessed December 2016.

[127] H. Bevrani, A. Ghosh, and G. Ledwich, "Renewable energy sources and frequency regulation: survey and new perspectives," *IET Renewable Power Generation*, vol. 4, no. 5, pp. 438–457, September 2010.

[128] J. Aho, A. Buckspan, J. Laks, P. Fleming, Y. Jeong, F. Dunne, M. Churchfield, L. Pao, and K. Johnson, "A tutorial of wind turbine control for supporting grid frequency through active power control," in *2012 American Control Conference (ACC)*, June 2012, pp. 3120–3131.

[129] W. Miller, Nicholas, M. Shao, and S. Venkataraman, *California ISO Frequency Response Study*, G.E. Energy, November 2011.

[130] N.-K. C. Nair and W. A. Qureshi, *Fault Ride-Through Criteria Development*. Singapore: Springer Singapore, 2014, pp. 41–67. [Online]. Available: http://dx.doi.org/10.1007/978-981-4585-27-9_3

[131] *Wind Farm Transmission Grid Code Provisions*, Commission for Energy Regulation, July 2004. [Online]. Available: http://www.cer.ie/docs/000178/cer04237.pdf

[132] F. Daz-Gonzlez, M. Hau, A. Sumper, and O. Gomis-Bellmunt, "Participation of wind power plants in system frequency control: Review of grid code requirements and control methods," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 551 – 564, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032114002019

[133] Z. Wu, W. Gao, T. Gao, W. Yan, H. Zhang, S. Yan, and X. Wang, "State-of-the-art review on frequency response of wind power plants in power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 1, pp. 1–16, Jan 2018. [Online]. Available: https://doi.org/10.1007/s40565-017-0315-y

[134] *Automatic Under-Frequency Load Shedding*, North American Electric Reliability Corporation (NERC), standard PRC-006-1. [Online]. Available: http://www.nerc.com/files/prc-006-1.pdf

[135] J. Goldman, *Fully 84 Percent of Hackers Leverage Social Engineering in Cyber Attacks*, eSecurity Planet, February 2017, available https://www.esecurityplanet.com/hackers/fully-84-percent-of-hackers-leverage-social-engineering-in-attacks.html. Accessed December 2017.

[136] *Stay one step ahead with GMS600 monitoring systems for generator-circuit breakers*, ABB, January 2015, available http://www.abb.com/cawp/seitp202/c363fc68cd993098c1257dcc0048ec76.aspx. Accessed December 2017.

[137] *PowerWorld Simulator*, PowerWorld Corporation, accessed May 2017. [Online]. Available: https://www.powerworld.com/products/simulator/overview

[138] *IEEEG1 Governor Model*, IEEE, accessed May 2017. [Online]. Available: https://www.powerworld.com/WebHelp/Content/TransientModels_HTML/Governor\%20IEEEG1\%20and\%20IEEEG1_GE.htm

[139] P. Pourbeik, *WECC Type 3 Wind Turbine Generator Model–Phase II*, Electric Power Research Institute, Jan. 2014. [Online]. Available: https://www.wecc.biz/_layouts/15/WopiFrame.aspx?sourcedoc=/Reliability/WECC%20Type%203%20Wind%20Turbine%20Generator%20Model%20-%20Phase%20II%20012314.pdf&action=default&DefaultItemOpen=1

[140] *WSCC 9 Bus System*, Western Electricity Coordinating Council, accessed May 2017. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/wscc-9-bus-system/

[141] B. Jairaj, *Indias Blackouts Highlight Need for Electricity Governance Reform*, World Resources Institute, August 2012, available http://www.wri.org/blog/2012/08/india%E2%80%99s-blackouts-highlight-need-electricity-governance-reform. Accessed December 2017.

[142] C. Barreto, J. Giraldo, . A. Cárdenas, E. Mojica-Nava, and N. Quijano, "Control systems for the power grid and their resiliency to attacks," *IEEE Security Privacy*, vol. 12, no. 6, pp. 15–23, Nov 2014.

[143] M. Q. Ali, R. Yousefian, E. Al-Shaer, S. Kamalasadan, and Q. Zhu, "Two-tier data-driven intrusion detection for automatic generation control in smart grid," in *2014 IEEE Conference on Communications and Network Security*, Oct 2014, pp. 292–300.

[144] R. Tan, H. H. Nguyen, E. Y. S. Foo, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, "Modeling and mitigating impact of false data injection attacks on automatic generation control," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1609–1624, July 2017.

[145] R. Tan, H. H. Nguyen, E. Y. S. Foo, X. Dong, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, "Optimal false data injection attack against automatic generation control in power grids," in *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, April 2016, pp. 1–10.

[146] Q. D. Vu, R. Tan, and D. K. Y. Yau, "On applying fault detectors against false data injection attacks in cyber-physical control systems," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, April 2016, pp. 1–9.

[147] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, ser. CCS '09. New York, NY, USA: ACM, 2009. [Online]. Available: http://doi.acm.org/10.1145/1653662.1653666 pp. 21–32.

[148] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 13:1–13:33, June 2011. [Online]. Available: http://doi.acm.org/10.1145/1952982.1952995

[149] M. A. Rahman, E. Al-Shaer, and R. G. Kavasseri, "A formal model for verifying the impact of stealthy attacks on optimal power flow in power grids," in *ICCPS '14: ACM/IEEE 5th International Conference on Cyber-Physical Systems (with CPS Week 2014)*, ser. ICCPS '14. Washington, DC, USA: IEEE Computer Society, 2014. [Online]. Available: https://doi.org/10.1109/ICCPS.2014.6843721 pp. 175–186.

[150] M. A. Rahman, E. Al-Shaer, and R. Kavasseri, "Impact analysis of topology poisoning attacks on economic operation of the smart power grid," in *2014 IEEE 34th International Conference on Distributed Computing Systems*, June 2014, pp. 649–659.

[151] M. A. Rahman, E. Al-Shaer, and R. B. Bobba, "Moving target defense for hardening the security of the power system state estimation," in *Proceedings of the First ACM Workshop on Moving Target Defense*, ser. MTD '14. New York, NY, USA: ACM, 2014. [Online]. Available: http://doi.acm.org/10.1145/2663474.2663482 pp. 59–68.

[152] M. A. Rahman, E. A. Shaer, and R. G. Kavasseri, "Security threat analytics and countermeasure synthesis for power system state estimation," in *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, June 2014, pp. 156–167.

[153] R. Tan, V. Badrinath Krishna, D. K. Y. Yau, and Z. Kalbarczyk, "Integrity attacks on real-time pricing in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 18, no. 2, pp. 5:1–5:33, July 2015. [Online]. Available: http://doi.acm.org/10.1145/2790298

[154] J. Giraldo, A. Cárdenas, and N. Quijano, "Integrity attacks on real-time pricing in smart grids: Impact and countermeasures," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–9, 2016.

[155] Y. Zhang, Y. Xiang, and L. Wang, "Power system reliability assessment incorporating cyber attacks against wind farm energy management systems," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–15, 2016.

[156] E. Ela, V. Gevorgian, P. Fleming, Y. Zhang, M. Singh, E. Muljadi, A. Scholbrook, J. Aho, A. Buckspan, L. Pao, V. Singhvi, A. Tuohy, P. Pourbeik, D. Brooks, and N. Bhatt, "Active power controls from wind power: Bridging the gaps," no. NREL/TP-5D00-60574, January 2014, National Renewable Energy Laboratory.

[157] J. Undrill, "Power and frequency control as it relates to wind-powered generation," no. LBNL-4143E, December 2010, Lawrence Berkley National Laboratory.

[158] Electric Ireland, "Valuesaver nightsaver electricity price plan," November 2015, Electric Ireland. [Online]. Available: https://www.electricireland.ie/switchchange/detailsValueSaverNightSaver.htm

[159] D. Wogan, *Electric utilities can now adjust your Nest thermostat to shift energy demand*, Scientific American, April 2013, https://blogs.scientificamerican.com/plugged-in/electric-utilities-can-now-adjust-your-nest-thermostat-to-shift-energy-demand/t.

[160] R. Tan, V. B. Krishna, D. K. Y. Yau, and Z. Kalbarczyk, "Integrity attacks on real-time pricing in electric power grids," *ACM Transactions on Information and Systems Security*, vol. 18, no. 2, pp. 5:1–5:33, July 2015. [Online]. Available: http://doi.acm.org/10.1145/2790298

[161] S. Fleten and E. Pettersen, "Constructing bidding curves for a price-taking retailer in the norwegian electricity market," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 701–708, 2005.

[162] M. Filippini, "Short-and long-run time-of-use price elasticities in Swiss residential electricity demand," *Energy Policy*, vol. 39, no. 10, pp. 5811–5817, 2011.

[163] M. G. Lijesen, "The real-time price elasticity of electricity," *Energy economics*, vol. 29, no. 2, pp. 249–258, 2007.

[164] M. Roozbehani, M. Dahleh, and S. Mitter, "Volatility of power grids under real-time pricing," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1926–1940, 2012.

[165] Schneider Electric, "Technical note of ION smart meter," www.bit.ly/15W9GR4.

[166] T. Nighswander, B. Ledvina, J. Diamond, R. Brumley, and D. Brumley, "GPS software attacks," in *CCS*, 2012.

[167] D. P. Chassin, K. Schneider, and C. Gerkensmeyer, "Gridlab-d: An open-source power systems modeling and simulation environment," in *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, April 2008, pp. 1–5.

[168] K. P. Schneider, Y. Chen, D. P. Chassin, R. Pratt, D. Engel, and S. Thompson, "Modern grid initiative distribution taxonomy final report," Pacific Northwest National Laborary, Tech. Rep., 2008.

[169] Australian Energy Market Operator, "2011 national electricity forecasting," www.bit.ly/ZQzkOs.

[170] Power Hero, "How energy is used in the home," https://www.powerhero.com.au/how-energy-is-used-in-the-home/.

[171] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 382–390, 2011.

[172] J. Lin, W. Yu, X. Yang, G. Xu, and W. Zhao, "On false data injection attacks against distributed energy routing in smart grid," in *ICCPS*, 2012.

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

[174] L. Jia, R. J. Thomas, and L. Tong, "Impacts of malicious data on real-time price of electricity market operations," in *Intl. Conf. System Sciences*, 2012.

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