Towson University
Office of Graduate Studies

TOWARDS A FRAMEWORK OF ENABLING EFFICIENT AND SECURED ENERGY BASED CYBER-PHYSICAL SYSTEM (CPS)

by
Guobin Xu

A Dissertation
Presented to the Faculty of the Graduate School
of Towson University
in Partial Fulfillment of the Requirements for the Degree of
DOCTOR OF SCIENCE
Department of Computer and Information Sciences

TOWSON UNIVERSITY
Towson, Maryland, 21252

July 2015

DISSERTATION APPROVAL PAGE

This is to certify that the dissertation prepared by Guobin Xu entitled Towards A Framework of Enabling Efficient and Secured Energy Based Cyber-physical System (CPS) has been approved by the thesis committee as satisfactorily completing the dissertation requirements for the degree Doctor of Science in Information Technology.

7/15/2015

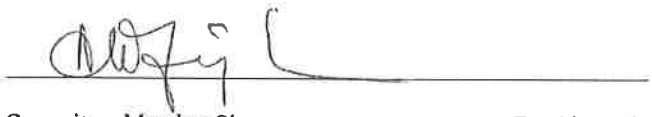
Chairperson, Dissertation Committee Signature Dr. Wei Yu Date

7/15/2015

Committee Member Signature Dr. Chao Lu Date

7/15/2015

Committee Member Signature Dr. Michael McGuire Date



7/15/2015

Committee Member Signature Dr. Alexander Wijesinha Date

7/28/15

Dean of Graduate Studies Dr. Janet V. DeLany Date

Dedicated

To

My Parents

Table of Contents

Abstract

Towards A Framework of Enabling Efficient and Secured Energy Based Cyber-physical System (CPS)

Guobin Xu

CPSs are referred to the systems that are built from the synergy of computational, communication, and physical components. The design of CPS tends to integrate computation and communication capabilities with monitoring and controlling of entities in the physical world. Unlike traditional embedded systems, CPSs are engineered physical systems, which are integrated, monitored and controlled by an intelligent computational core. A number of CPSs, including the Smart Grid, process control systems, transportation systems, and healthcare systems are expected to be developed using advanced computing and communication technologies.

The Smart Grid, as an energy-based CPS, must be dependable, cost-effective, secure, safe, and efficient and operate in real-time, which is a highly distributed and complicated system that manages and controls geographically dispersed assets. This kind of system inherently operates under the presence of uncertainty or unpredictable behavior due to intrinsic and extrinsic courses and cyber adversaries. By providing efficient and secured operation performance of the Smart Grid, high volume data streams associated with the Smart Grid operations need to be quickly

processed and analyzed, raising significant challenges, which can hinder the effectiveness of systems themselves.

To address those challenges and corresponding issues, in this dissertation, we develop a framework to enable efficient and secured energy based CPS by developing effective schemes to address the uncertainty in both cyber and physical components in the Smart Grid. To be specific, first, to adapt physical uncertainties, we develop techniques to effectively manage distributed renewable energy resources to make the energy transmission and distribution more efficient. Second, to investigate cyber uncertainties to the system, we systematically explore the space of attacks in energy management and investigate the risk of those attacks and countermeasures. Finally, we investigate cloud computing to efficiently store and process big data for the Smart Grid operations and security management.

Acknowledgement

I would like to express my sincere gratitude to my academic advisor, Dr. Wei Yu, who is a renowned researcher in the field of cyber security, networking, and cyber-physical systems areas. He frequently shares with me potential research topics and encourages me to explore different perspectives of thinking. I would have never been able to earn my doctorate degree without his guidance. Due to his motivation and advice, I believe I will have great success in academic research.

I would also like to thank Dr. Chao Lu, who encourages me to continue my high level study in Towson University. He encourages me to practice by myself and helps me to build confidence in my research.

I would also like to acknowledge Dr. Alexander Wijesinha and Dr. Michael McGuire for providing guidance on my dissertation. In addition, I would like to thank my colleagues and our research group members who collaborates with me and provided help in my academic endeavors.

I would like to thank my fiancee, Ms. Jin Guo, for standing by me and helping me during difficult times. Finally, I would like to thank my parents for continuous their support and their deep love.

Towson, Maryland Guobin Xu

July, 2015

List of Tables

Page

List of Figures

Page

Chapter 1

Introduction

1.1 Motivation

A typical cyber-physical system (CPS) is defined as a system that features a tight integration of computation, networking, and physical elements [1]. It covers the smart transportation, smart electrical power grid, smart medical systems, smart manufacturing systems, etc. In these systems, the geographically distributed sensors, actuators, and controllers are tightly integrated through communication networks and computation cores, enabling the secured and efficient operations of physical systems. Sensors obtain measurements from physical systems and transmit measured data to operation centers through communication networks that computation cores can be further used to determine the status of physical systems. The operation center then performs management and control, and sends control commands to actuators, which makes physical systems in desirable states. A CPS (e.g., Smart Grid [2]) is a highly distributed and complicated system that manages and controls geographically dispersed assets. The intrinsic complexity, heterogeneity, and sensitivity to system performance make modeling, simulation, emulation, and design of such a system very challenging.

The Smart Grid, as an energy-based CPS, uses modern communication technologies to achieve a more efficient, reliable, secure, and resilient power grid. In

the US and many other countries, the modernization of the electric power grid is vital to the national efforts in order to increase energy efficiency, transition to distributed renewable energy sources, reduce greenhouse gas emissions, and build a sustainable economy that ensures prosperity for current and future power generations [3,4]. To this end, the development of the Smart Grid has received a renewed attention in recent years [2]. With the Smart Grid, users can be provided with reliable supply of power regardless of their location and time. A high efficiency in power generation and resource utilization can also be realized via monitoring and controlling the power transmission and distribution process.

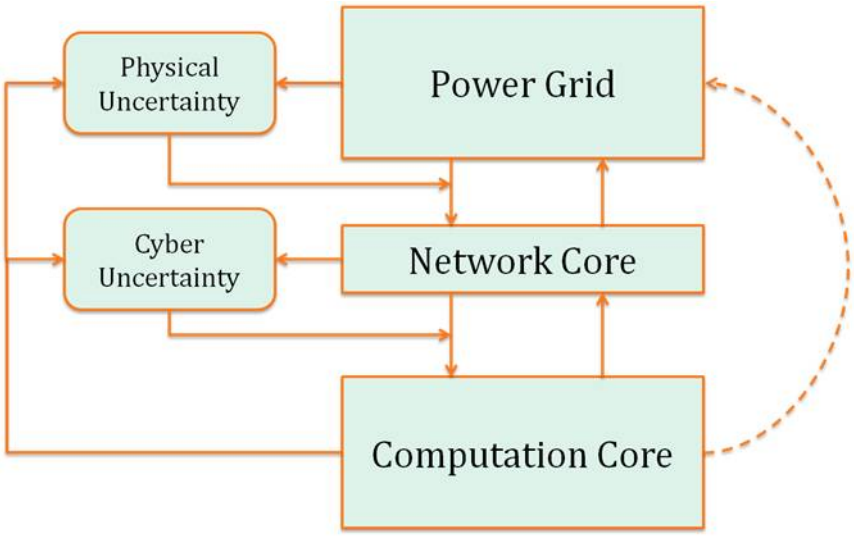


Figure 1.1: Coordination of Cyber and Physical Components

In the Smart Grid, as stated in Figure 1.1, the management of energy resources needs the coordination from both cyber and physical components. Nonetheless, the Smart Grid inherently operates under the presence of uncertainty or unpredictable

behavior due to intrinsic and extrinsic courses and cyber adversaries. For example, in communication and computation components, uncertainties can be raised by benign faults, attacks, congested networks, computing capacity, etc. Therefore, the faults/errors of computational elements, sensors and actuators, physical components and attacks on network components, applications, and communication cores, compose serious security problems, which have to be addressed. Uncertainties that could be randomly introduced or raised by malicious adversaries on purpose at either physical or cyber network components will affect the decision at computation/decision core.

Uncertainties are mixtures of two dimensions: $X : cyber, physical$ and $Y : failure, attack$. For example, the false data injection attack proposed in [5] against the state estimation of power grid belong to $cyber, attack$ category and the Hurricane Sandy in late October 2012 [6], cutting power in and around cities, belongs to $physical, attack$ category. It is critical to quantify natural uncertainties from physical components (e.g., solar irradiance, wind speed, temperature, and others), users (e.g., plug-in vehicles), and malicious adversaries (e.g., injecting false information to power grid).

In the computation core, by providing secured and efficient the Smart Grid operation performance, high volume data streams associated with system operations need to be quickly processed and analyzed, raising significant challenges, which

can hinder the effectiveness of systems themselves. In addition, efficiently processing threat monitoring data from both physical components and cyber components will facilitate the detection of cyber-threats as well as help security administrators respond to cyber-threats in a timely manner. The processing of a massive threat and monitoring data to generate accurate and timely security alerts is also challenging.

1.2 Significance of Proposed Research

In this dissertation, we propose to investigate a framework of enabling efficient and secured energy based CPS - Smart Grid. To do so, we make several contributions.

- First, to adapt uncertainties, we develop techniques to effectively manage energy resources. We develop schemes to model distributed energy resource (DER) that can be used to replace traditional fossil fuel resources (e.g., coal, oil, and others) [7]. We also model the constraints of DER and integrate those constraints into power generation dispatch schemes and evaluate their impact on bulk power generation. We consider the reliability constraint of DER and use reliability theory and define metrics to quantify the generation capacity of DER. Based on the real-world power-generation data and load-demand data, we investigate the probability of reliable power generation by wind and solar energy during a period of time. We use different size of IEEE power grid bus systems to evaluate the effect on the bulk power generation in terms of

different scaled levels of DER.

- Second, to investigate malicious uncertainty to the system, we systematically explore the space of attacks in the energy management process, including modules being attacks (e.g., end-user, communication network, system operations, etc.), attack avenue (e.g., confidentiality, integrity, availability, etc.), attack strength (e.g., strong, stealthy), and system knowledge (e.g., full, partial). Specifically, we take the attacks against distributed energy transmission as a case study to investigate the risk of those attacks. We also develop the defense taxonomy to secure energy management with three orthogonal dimensions: methodology, sources, and domains.

- Third, we investigate a cloud computing based architecture to assist assist efficient and secured Smart Grid operations, which offers vast storage, flexible deployment, more computation resources, less expensive infrastructure investment, and ubiquitous sharing of information across all members of the cloud. Our proposed system architecture consists of three main components: data sources, cloud infrastructure, and an operation center. We leverage the cloud infrastructure to develop a cost-effective data stream storage. By leveraging MapReduce-based data processing, we improve data storage efficiency, speed up access to data, and eliminate operational delays. Through extensive experiments, our data shows that our developed system can efficiently process

and analyze a large amount of data.

1.3 Organization of Dissertation Research

This dissertation is structured as follows. In Chapter 2, we introduce the background of CPS and Smart Grid. In Chapter 3, we conduct literature review of general CPS, Smart Grid, and cloud computing. In Chapter 4, we present the efficiency energy resource management in the Smart Grid by investigating the effectiveness of integrating distributed energy resources. In Chapter 5, we discuss secured energy resource management in the Smart Grid and explore the space of threats in the energy management process. In Chapter 6, we investigate cloud computing based techniques to assist secured and efficient Smart Grid operations. Finally, we conclude the dissertation in chapter 7.

Chapter 2

Background

2.1 Cyber-physical System (CPS)

Generally speaking, a CPS is a large distributed and complex system, featuring a tight integration between cyber computation, communication, and physical components. To enable effective integration, a number of geographically dispersed components need to be effectively managed and controlled. Nonetheless, the intrinsic complexity, heterogeneity, and sensitivity to the system performance make the modeling and design of a CPS challenging. For example, the space CPS is a typical instance of a CPS, which tends to effectively provide a global coverage for supporting heterogeneous mobile devices (e.g., people, vehicles, and manned/unmanned aircraft, etc.) and delivering services (e.g., voice, video, and data transmission) to end users.

To address integrating cyber and physical components, a generic layered framework for CPSs is shown in Figure 2.1. Our proposed layered framework consists of four layers: physical layer, sensor layer, network layer, and service layer.

- *Physical Layer:* There are hardware devices composed of electronics, actuators, and memory states. These elements are combined for computation, processing, and various functions, which can be viewed as agents of sensors.

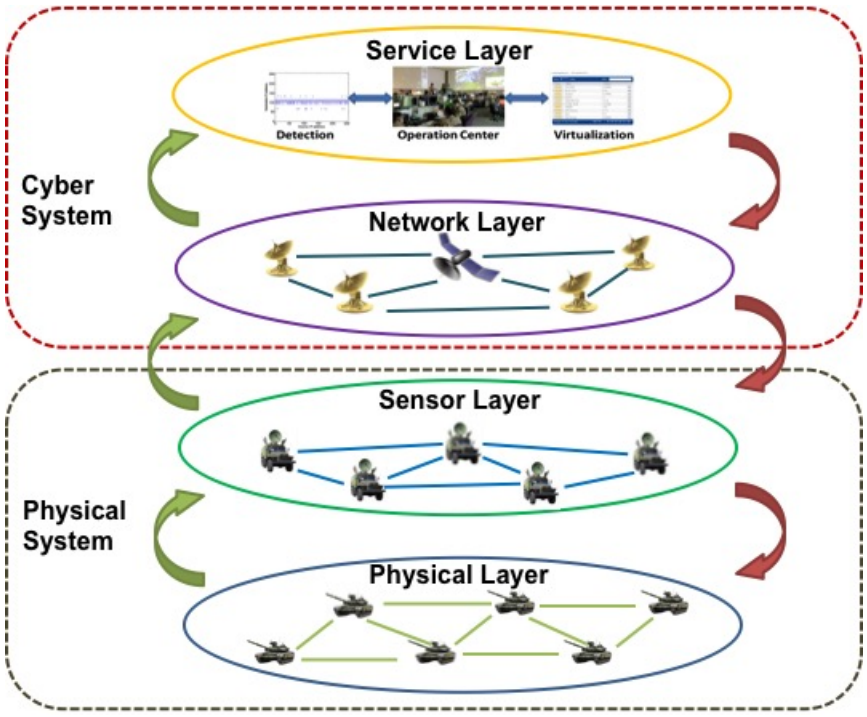


Figure 2.1: A Layered Architecture for CPSs

- *Sensor Layer:* Sensors are connected through cyber communication networks to measure the characteristics associated with the physical components. Because sensors may be deployed in unattended or even harsh environments, the lack of tamper-resistant hardware increases the possibility of nodes being compromised by cyber adversaries. Due to the limited resources, both lossless and lossy aggregation techniques can be used in sensor networks to reduce the resource usage (e.g., energy consumption and bandwidth) for transmitting information through the network while preserving the desired accuracy for CPS operations [8].

- *Network Layer:* The nodes can establish self-organizing networks. For a large CPS, the nodes can self-organize as a cluster in a lower level. Each cluster has a cluster-header master, which can aggregate activities in the cluster and some cluster-header nodes can organize as a cluster in a higher level. In each cluster, sensors can directly communicate with each other in the same cluster. The cluster-header node with the highest level can directly communicate with the operation center.

- *Service Layer:* The operation center is responsible for coordinating nodes distributed over a CPS and carrying out a CPS situational awareness. The operation center will receive and analyze the measurement information from the nodes in a CPS. The operation center would also provide tools to enable analysis methods to monitor, visualize, and query the status of the secured operations within a CPS.

2.2 Smart Grid

The Smart Grid uses modern advanced cyber computing and communication technologies to achieve more efficient, reliable, secure, and resilient operations and services as compared to existing legacy power grid. The development of power system and information communication technologies fosters the development of efficient system operations.

Energy management, including the distributed energy generation, transmission, storage, and distribution, will change the way, which we consume and produce energy in a similar way like the people using Internet [9]. The energy management in the Smart Grid is critical in addressing the challenges of meeting the world's growing energy demand by integrating various renewable energy resources. For example, a user may collect and store the energy resource and then push them to the grid or pull them from the grid through power connection interfaces. The energy management will provide the intelligence to the grid, situational awareness and dynamic optimization of operations to enhance the reliability of the power system, power quality, efficiency, scalability, self-healing capability and security of power system. It enables the reduction of outage occurrences and durations, the minimization of losses through monitoring, the optimization of utilizing assets by management of demand and distributed generation, and the reduction of maintenance costs. Nonetheless, to develop an effective energy management process in the Smart Grid, there are two unique challenges as follows.

- First, the Smart Grid is a highly distributed and complicated system that manages and controls geographically dispersed assets, often scattered over thousands of square kilometers. The inherent complexity and heterogeneity, and sensitivity to timing pose modeling and design challenges. To achieve flexible implementation and interoperability, National Institute of Standards

and Technology (NIST) developed the reference model that divides the Smart Grid into seven domains: customers, markets, service providers, operations, bulk generation, transmission, and distribution [3]. Each domain and its subdomains encompass the Smart Grid actors and applications. These actors and applications, as well as many factors from the grid, network, and users have impact on the system performance. Unfortunately, there is lack of a systematical approach to model and analyze the energy management with the consideration of diversified actors and applications, including energy demanders, energy supply resources, and the ability of integrating both cyber and physical components and reflecting the interactions between them.

- Second, the Smart Grid inherently operates under the presence of uncertainties or unpredictable behavior due to intrinsic and extrinsic environments. In particular, in cyber communication and computational components, uncertainties can be benign faults, malicious attacks, network congestion, delay, time synchronization, and computational capacity. In the physical component, uncertainties can be from either nature or human. On the natural side, widely distributed solar irradiance, wind speed, temperature, and humidity will introduce the great uncertainty of distributed energy resources. This leads to the energy supply uncertainty. The natural disaster may cause blackouts over large regions, tripped by minor events with surprising speed into widespread

power failure. Demanders are uncertain over the course of the day, extreme weather events, and the seasons of the year, which will lead to the energy demand instability. Therefore, the identification and treatment of both the benign faults and malicious attacks in both cyber and physical components must be developed in order to achieve a useful and effective Smart Grid.

Chapter 3

Literature Review

In this section, we briefly review the related work. Given the multi-disciplinary agenda of our proposed research, we only cover the most related work in this section.

3.1 General CPS

The design of CPS tends to integrate computing and communication capabilities with monitoring and control of entities in the physical world. Unlike more traditional embedded systems, CPS is natural and engineered physical systems, which are integrated, monitored and controlled by an intelligent computational core [10]. Lee *et al.* [11] proposed to establish a new development paradigm, which enables the effective design, implementation, and certification of medical device CPS. Since the initiation of CPS from the US National Science Foundation (NSF) in 2006, researchers from different fields discussed the related concepts, technologies, applications, and challenges during CPS Weeks [12] in the past few years and international conference on the CPS subjects [13, 14]. There are some literatures related to energy issues in CPS [15, 16]. For example, Rajkumar *et al.* [17, 18] presented a future vision of CPS and identified some specific CPS applications, including advanced electric power grid system and grand challenges. Wan *et al.* [19] proposed

a framework to analyze the design of CPS using the in-home health care system as an example. In CPS 08 summit [20], there are several research efforts, which described the grand visions and grand challenges used to drive research in CPS. In this summit, the future distributed energy systems such as the Smart Grid were identified as an important CPS application.

Security is a critical issue for CPSs and a number of research efforts have investigated the security of CPSs [21]. For example, the challenges in the security of CPSs have been partially addressed in [22, 23]. Shafi [24] conducted a review of existing research efforts on CPS security. Zhu *et al.* [25] described a layered architecture towards a secure CPS, which helps to identify research problems and challenges in each layer and to build models for designing security measures for the control systems of critical infrastructures. The CPS security problem is comprised of CPS detection of element faults and errors, assaults on cyber communication networks (e.g., operating systems, applications, data, etc.), and attacks against the communication between cyber components and physical components. The available techniques only consider or address a part of the CPS security problem. For example, in [26], the communication between cyber components and physical components is considered as ideal (i.e., neither attacks nor faults). Ground tracking control systems in the satellite based space system is an example of a typical CPS [27–30] as well. In order to take advantage of CPS to support key elements

of future missions, the Ames Center Chief Technologist (CCT) Office has established the Cyber-Physical Systems Modeling and Analysis (CPSMA) to apply CPS technologies to specific requirements for long-duration human spaceflight [31]. In addition, the Idaho National Laboratory published control Systems Cyber Security: Defense in Depth Strategies [32], where a number of cyber security issues, including cyber attacks and wireless attacks, were presented.

To secure CPSs, game theory, resilient control, attack detection, forensic techniques, and others, have been studied [26, 33–39]. For example, Li *et al.* [33] presented a game-theoretic method for sensor attack avoidance. Pasqualetti *et al.* [34] proposed a unified framework to analyze the resilience of CPSs against threats. Cardenas *et al.* [35] presented a game theoretical model to defend against bad data and protect the privacy of users in the Smart Grid. Chris *et al.* [36] presented a number of game theoretic formulations of attack and defense aspects in CPSs with different cost and utility functions for the attacker and defender. Zhu *et al.* [37] presented a game theoretic based resilient control framework to deal with attacks on cyber components and faults of actuators with perfect sensors and communication. Jin *et al.* [26] applied resilient control techniques in CPSs to deal with component failures. Mitchell *et al.* [38] analyzed the effects of intrusion detection and response on CPSs comprising sensors, actuators, control units, and physical objects for controlling and protecting physical infrastructures.

3.2 Smart Grid

The development of the Smart Grid has received a renewed attention recently [2,4] and is an energy-based CPS [40]. The Smart Grid specifies additional intelligence and bidirectional communication and energy flows to address the efficiency, stability, and flexibility that plague the grid [3, 18]. McMillin *et al.* [41] defined the relationship of power-grid system with the CPS and stated the power grid, power electronics, and embedded control software form a CPS, whose design is heavily influenced by fault tolerance, security, decentralized control, and economic/ethical social aspects. The Smart Grid provides services include a wide-scale integration of renewable energy sources, provision of real-time pricing information to consumer, demand response programs involving residential and commercial customers, rapid outage detection, and granular system health measurement. To enable the Smart Grid, a number of research efforts have been paid in distributed energy management [42,43]. For example, Xie *et al.* [44] investigated a novel modeling paradigm, which seamlessly integrates physics-based and data-driven models of distributed resources for provision of energy storage services in power systems. A number of research efforts have been made to study factors, requirements, and the effects of integrating distributed energy resources and energy storage technologies into the grid [45–53]. For example, Taneja *et al.* [48] obtained the measurement of the real-time blend of supplies on the primary California grid and scaled the solar and wind

assents to a level of penetration. Ilic *et al.* [49] proposed a novel look-ahead interactive dispatch, which can integrate wind power, price-responsive demand and other distributed energy resource, internalizes inter-temporal constraints at distributed energy resources level, and dispatches the results of distributed decisions subject to spatial security constraints.

Security is one of the active research areas in the Smart Grid [54–56]. For example, NIST [3] highlighted the cyber security challenges in the Smart Grid. Patrick *et al.* [57] discussed the risks of the security and privacy issues. Adam *et al.* [58] introduced an access graph based approach to determine the system's attack exposure and evaluate the security exposure of large scale the Smart Grid environment. False data injection attacks against the state estimation of power system were deeply investigated [5, 59]. For example, Liu *et al.* [5] proposed that the adversary could manipulate the state estimate and bypass the traditional bad data detection and identification algorithms, with the knowledge of network configuration. Stephen *et al.* [60] developed an archetypal attack tree approach to model the broad adversary's goals and attack vectors and guide the evaluators to use grafted tree as a road map to penetration testing. Lin *et al.* [61] studied the vulnerability of distributed energy routing process and investigate novel false data injection attacks against the energy routing process.

3.3 Cloud Computing

Generally speaking, cloud computing is a technology that uses the Internet and central remote servers to provide computation, software, data access, and storage services, which do not require end-user knowledge of the physical location and configuration of the servers. Most current clouds are built on the top of a modern data center [62]. Cloud incorporates different service models [63] such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) and so on. Cloud computing is gaining popularity in both academia and industry and both have been active in the research on cloud computing architectures. For example, Vecchiola *et al.* [64] introduced a .NET-based Cloud Computing platform, which provides a set of APIs that allow developers to build .NET applications that leverage their computation using the cloud. Huang *et al.* [65] developed a low-cost, scalable, and secured platform that enables web-delivery of application-based services with a set of common business and operational services.

To provide efficient data storage in the cloud, a number of research efforts have been performed [66–72]. For example, Zeng *et al.* [66] proposed a layered and cooperative architecture of the cloud storage system. Abu-Libdeh *et al.* [67] proposed to apply RAID (Redundant Array of Inexpensive Disks) techniques to store data across multiple providers, which can reduce the storage switching costs and achieve a high resilience against outages and failures. Kamara *et al.* [68] proposed a

combination of recent and non-standard cryptographic primitives to build a secure cloud storage on top of a public cloud infrastructure. Harnik *et al.* [69] investigated the side channels in a cloud environment. Wu *et al.* [70] provided a comprehensive review of the key technologies in cloud computing and cloud storage, the types of cloud services, and describes the advantages and challenges of cloud storage. Wang *et al.* [71] proposed a scheme for confidential data sharing on cloud servers.

To conduct big data analysis, there are a number of research efforts using the MapReduce framework [73–87]. For example, Jeffrey *et al.* [73] introduced the *MapReduce* programming model and demonstrated how the *MapReduce* works with the implementation in a large cluster. Jimmy *et al.* [74] investigated the *MapReduce* algorithm design with a focus on text processing (e.g., natural language processing, information retrieval, and machine learning). Morken *et al.* [75] studied the *MapReduce* model and the frameworks for netflow data processing. Ebrahimi *et al.* [76] investigated how to solve linear programs using *MapReduce*. David *et al.* [77] investigated an elastic streaming *MapReduce* for distributed data stream processing. Felix *et al.* [79] designed and implemented a *MapReduce* based max-flow algorithm to process large small-world graphs. Osterman *et al.* [80] investigated the potential and the usability of cloud computing for the scientific community. Gunarathne *et al.* [81] studied the implementation of the MapReduce in the cloud for science

applications. Shivhare *et al.* [83] investigated the pros and cons of cloud computing in storing and processing big data and discussed potential solutions. Chen *et al.* [84] evaluated a cloud based security center for traffic data forensic analysis and proposed a collaborative cyber security management system for data collection and analysis using a parallel programming paradigm in a cloud computing platform. Liu *et al.* [85] leveraged the MapReduce technology to build a highly-scalable system on top of a scalable cloud. Tan *et al.* [86] developed a tool namely Kahuna to diagnose performance problems in MapReduce systems. Zhang *et al.* [87] developed a simple Matlab-to-MapReduce translator for cloud computing, namely M2M, for the basic numerical computations capable of translating a Matlab code with up to 100 commands to a MapReduce code in just a few seconds.

As an open source software framework and an implementation of *MapReduce* algorithm, a number of research efforts have been conducted to use *Hadoop* to process large scale data sets on large clusters [88–96]. For example, Tom [89] introduced and demonstrated how to process big data using *Hadoop*. Dhruba *et al.* [90] introduced the distributed file system architecture and designs in *Hadoop*. Shvachko *et al.* [92] described the architecture of a Hadoop File system and provided an implementation report on managing 25 petabytes of enterprise data. Kambatla *et al.* [93] investigated the performance of existing solutions in optimally provisioning

the MapReduce job in the Hadoop File system environment. Xie *et al.* [94] proposed a data placement scheme, which balances data across nodes before processing the load on a heterogeneous Hadoop MapReduce cluster. Sandholm *et al.* [95] proposed a dynamic priority parallel task scheduler for Hadoop. Their scheduler enables the user to control their allocated capacity and dynamically adjust their spending based of the current demand for cloud services. Shafer *et al.* [96] investigated the root causes of Hadoop performance bottleneck and discussed the tradeoffs between portability and performance in the Hadoop distributed file system.

Chapter 4

Efficient Energy Resource Management in The Smart Grid

4.1 Overview

The Smart Grid is the integration of cyber communication networks with the physical power grid in order to create two-way communication in power systems, extending from power generation nodes through transmission and distribution into end user premises [3]. With the Smart Grid, utility companies could have near-real-time information to manage the entire power grid by actively sensing and responding to changes in energy demands, supplies, and costs [97]. Generally speaking, the Smart Grid is stated as a fundamental re-engineering of the electricity service industry and a modernization of the world's electrical grids.

In the recent past, a massive power failure hit India for a second day running, leaving more than half the country without power affecting over 600 million people. The failure was caused by three regional power grids overdrawing power from the grid to meet heavy demand [98, 99]. The ever-increasing demand on power usage brings more concern on stable power supply, which fosters the integration of distributed renewable energy resources and energy storage technologies into the grid.

Distributed energy resources and energy storage devices are certain to have a significant impact on the cost efficiency and service reliability of the power grid.

Because distributed energy resources have time-dependent and uncertain factors, it is hard to predict their power generation. Consequently, integrating distributed energy resources into the traditional power grid will incur service reliability issues, which needs to be carefully addressed. Therefore, there is a urgent need for the research, aimed at understanding and quantifying the impact that the massive integration of distributed energy resources will have on the power system in terms of efficiency, operational reliability of the power grid and economic consequences [100].

In this chapter, as an example of efficiency energy resource management in the Smart Grid, we investigate the effectiveness of integrating distributed energy resources in the Smart Grid. To be specific, we first introduce three types of power grid systems: traditional bulk power grid, power grid integrated with distributed energy resources, and power grid integrated with both distributed energy resources and storage device. We then model and analyze the effectiveness of integrating distributed energy resources and energy storage devices in the Smart Grid based on the metrics of power cumulative cost and user service reliability. Notice that the power cumulative cost is defined as the cumulative cost of the power generation and distribution processes in a time interval and the user service reliability is defined as the probability that the quantity of user requested power can be met.

Based on the California Independent System Operator Corporation (ISO) data [101], we conduct simulations using Matlab simulation tool. Our evaluation results

confirm that the integration of distributed energy resources could increase the grid service reliability to meet customer demands and reduce the power generation and distribution cost, and the integration of energy storage devices could effectively smooth the bulk power generation, mitigate the grid peak load, and reduce the impact from uncertainty. As examples, in our experimental settings, we can see that the power cost in generation and distribution process could save about $40,000$\$ after integrating distributed energy resources. With integrating a 500 MW capacity energy storage device, the bulk power generation can be stable at 2200 MW that indicates the storage device could reduce the disturbance from uncertainties.

4.2 Power Grid Systems

In this section, we first give the overview of the Smart Grid. Then, we introduce three types of power grid systems.

4.2.1 Overview

The Smart Grid is the integration of cyber computer and communication technology into the power grid, where each of the key components within the grid is interconnected through a communication network. Its goal is to achieve sustainable integration and utilization of renewable energy resources, energy efficiency, security, reliability, reduction of greenhouse gas emissions, and a sustainable economy that ensures prosperity for current and future generations [3].

The Smart Grid is self-healing, self-balancing and self-optimizing. The grid failures can be automatically detected or even predicted based on the real-time data analysis. With the advanced metering infrastructure, the two way communication between the user and utility can be facilitated. With the collected information, the power demands from users can be accurately forecasted to adapt the power generation at utility, which could ultimately reduce the power generation cost and optimize power production.

To reduce the greenhouse gas emission, distributed energy resources will be widely deployed and used throughout the Smart Grid, which could generate power instead of relying on fossil fuel as the traditional energy resources. Nonetheless, how to effectively integrate distributed energy resources (e.g., wind energy, solar energy, and others) to the grid and have controllable impact on the grid is crucial for the development of the Smart Grid. To this end, in the following, we introduce three types of power grid systems.

4.2.2 Traditional Bulk Power Grid

In the traditional bulk power grid, the energy generation and distribution processes are simple. Users can only obtain the power supply from the bulk power generator. With lots of uncertainties from user demand, the service reliability in the traditional bulk power grid is insured through over-provisioned power and excess capacity. That is, power generators should produce more power than user demand.

Nonetheless, the ever-increasing demand on energy usage has led to a serious depletion of fossil-fuel reserves. Therefore, it is likely that the grid will not have sufficient capacity to meet future demand. According to the North American Electric Reliability Corporation, forecast demands for electricity may exceed projected available capacity in the United States by 2015 [102]. Traditional bulk power grid is an inefficient and environmentally wasteful system. For example, traditional bulk power generation causes 25.9% of global carbon (CO2) emissions and CO2 emissions [103] from electricity use will grow faster than those from all other sectors through 2050 [104]. Conclusively, traditional power grid is becoming more fragile, less reliable, and less efficient.

4.2.3 Power Grid With Distributed Energy Resources

Distributed energy resources such as wind turbines and solar plants are intermittent by nature and thereby unpredictable. A rapid or unpredicted change in weather such as the variability in wind speed could cause error in power generation forecast and seriously limit the capacity of distributed energy resources, posing negative impact on the grid. Despite these uncertainties, power grid system operator shall balance user demands and available power supply with appropriate generation resources in real-time.

To effectively integrate distributed energy resources, the power system operator shall estimate the status of distributed energy resources, which can control

whether distributed energy resources should be integrated to the grid. To measure the impact of integrating distributed energy resources and the efficiency of system operation, we define two metrics as following:

Definition 1: The Power Cumulative Cost is defined as the cumulative cost of the power generation and distribution processes in a time interval.

The power generation cost can be computed by the marginal cost of power generation (denoted as C_i) and the quantity of power that generators can supply (denoted as $G_{ij}(t)$). The power distribution cost can be computed by marginal cost of the power loss (denoted as C_p) and the quantity of power loss in the distribution process (denoted as $\Delta G_{ij}(t)$). Notice that i and j are denoted as the ID of the generator and user, respectively.

Definition 2: The User Service Reliability is defined as the probability that the quantity of user requested power can be satisfied, which can be represented as $SR(t)$. The detailed analysis of these two metrics can be found in Section 4.3.

Unfortunately, integrating renewable energy resources is not an ultimate solution. The intermittent nature of renewable energy resources and the mismatch of generated power with peak load conditions pose high pressing challenges to the effective integration of renewable energy resources into the grid. Explicitly, renewable energy resources such as wind turbines and solar plants are intermittent by nature and unpredictable as they fluctuate randomly and mostly beyond human

control.

4.2.4 Power Grid With Both Distributed Energy Resources and Energy Storage Devices

To smooth the bulk power generation and mitigate the intermittency and uncertainties of distributed energy resources, energy storage devices should be integrated to the grid. Generally speaking, standing as a buffer between end users and power generators, energy storage devices could store power during times when generation from bulk power generators exceeds consumption. The stored power could be used at times when consumption exceeds generation. Otherwise, in traditional bulk power grid, power generation needs to be drastically scaled up and down to meet momentary consumption.

By integrating energy storage devices, bulk power generation could be maintained at a constant level. In addition, the energy storage will provide backup power and provide uninterrupted power, which could mitigate power fluctuations and uncertainties of power generators. For example, grid energy storage devices (e.g., battery and plug-in hybrid electric vehicles (PHEVs)) could generate and consume electric power intelligently, reducing the disturbance caused by peak and non-peak power usage, and power generation costs.

4.3 Modeling and Analysis

In this section, we model and analyze the effectiveness of the three systems based on the metrics of cumulative power cost and user service reliability.

4.3.1 Cumulative Power Cost

4.3.1.1 Traditional Bulk Power Grid

Recall that in Section 4.2.2, in the traditional bulk power grid, we consider a number of nodes, which request power from the bulk power generation through energy transmission links. Power loss is the main issue for energy distribution process, which can be denoted as $\Delta G_{ij}(t)$. When a current flows through the power line, some of the power is converted to electric heating loss resulted from transmission resistance. The power loss is mainly caused by power line heating which can be denoted as $\Delta G_{ij}(t) = I_{ij}^2(t) \cdot R_{ij}$. With $R_{ij} = \rho \frac{l_{ij}}{S}, \quad \forall i \in G, \quad \forall j \in D$ and $G_{ij}(t) = I_{ij}(t) \cdot U_{ij}(t)$, the power loss can be represented as $\Delta G_{ij}(t) = k \cdot G_{ij}^2(t) l_{ij}, \quad \forall i \in G, \quad \forall j \in D$,

From this expression, we can infer that there are two options to reduce power loss: (i) reducing the power line resistance, and (ii) reducing the power line current. To reduce the resistance of a power line, we consider the resistance of the line as given by the laws of resistance and can be represented as $R_{ij} = \rho \frac{l_{ij}}{S}, \quad \forall i \in G, \quad \forall j \in D$, where ρ is the resistivity, and S is the line cross section. Because increasing the

line cross-sectional area is too costly, we assume all lines cross-sectional area is a constant value. Hence, the transmission distance is the key factor that determines power line resistance. To reduce current flow, with a certain demand power of $G_{ij}(t) = I_{ij}(t) \cdot U_{ij}(t)$, we need to increase the voltage in order to reduce the current flow, which is the reason of high-voltage transmission. To simplify our analysis, we consider that all the power lines have the same high voltage and we define a parameter k to represent $\frac{\rho}{S \cdot U_{ij}^2(t)}$. Then, the power loss can be represented as

$$\Delta G_{ij}(t) = k \cdot G_{ij}^2(t) l_{ij}, \quad \forall i \in G, \quad \forall j \in D, \tag{4.3.1}$$

where k is $\frac{\rho}{S \cdot U_{ij}^2(t)}$, l_{ij} is the line distance between generator i and user j, ρ is the resistivity, S is the line cross section, $G_{ij}(t)$ is power distribution from generator i to user j, $I_{ij}(t)$ is power current of the line l_{ij}, $U_{ij}(t)$ is power voltage of the line l_{ij}, G is the set of generators, and D is the set of users.

Hence, the grid power balance equation can be derived by,

$$G_{ij}(t) - \Delta G_{ij}(t) = D_{ij}(t), \tag{4.3.2}$$

where $G_{ij}(t)$ is the quantity of power that the generator i provides user j and $D_{ij}(t)$ is the quantity of power that the user j receives from generator i. Then, the cumulative power cost of the traditional bulk power grid can be represented as,

$$\begin{cases} C = C_i \cdot \sum\limits_{j \in D} G_j(t) + C_p \cdot \sum\limits_{j \in D} \Delta G_j(t), \\ G^{min} \leq \sum\limits_{j \in D} G_j(t) \leq G^{max}. \end{cases} \tag{4.3.3}$$

4.3.1.2 Integrating Distributed Energy Resources

Recall that in Section 4.2.3, to integrate distributed energy resources, we shall consider the following constraints: (i) The objective of energy distribution is to minimize the cumulative power cost; (ii) The generated power can not exceed the capacity of the power generator; (iii) The total power generation shall be equal to the total power demand plus the total power loss; (iv) Each power transmission line cannot exceed its capacity; (v) The demand power shall be equal to the sum of power transmitted from different generators; (vi) The power generated by the generator minus the power loss should not be less than the user demand power from the generator. Based on the above constraints, we formalize the energy distribution process as an non-linear optimization problem listed below.

$$\textbf{\textit{Objective.}} \quad Min\left\{C_i \cdot G_T(t) + C_p \cdot \Delta G_T(t)\right\} \tag{4.3.4}$$

S.t.

$$\begin{cases} \forall i \in G, \forall j \in D, \\ G_{ij}(t) \leq L_{ij}, \\ G_i^{min} \leq G_i(t) \leq G_i^{max}, \\ G_i(t) = \sum\limits_{j \in D} G_{ij}(t), \\ D_j(t) = \sum\limits_{i \in G} G_{ij}(t) - \sum\limits_{i \in G} \Delta G_{ij}, \\ \Delta G_{ij}(t) = k \cdot G_{ij}^2(t) l_{ij}, \\ G_T(t) = \sum\limits_{i \in G} G_i(t), \\ \Delta G_T(t) = \sum\limits_{i \in G} \sum\limits_{j \in D} \Delta G_{ij}(t). \end{cases}$$

Because distributed energy resources are renewable and with low cost, their generation marginal cost is less than that of bulk power generation. The decrease of power transmission distance and the quantity of power generation will result in the decrease of power loss. To meet the user demand, the power supplied from nearby generators will reduce the power transmission loss and reduce bulk power generation.

4.3.1.3 Integrating Both Distributed Energy Resources and Energy Storage

Suppose that a storage device that could be discharged or recharged from the grid in a time interval t. This storage device will discharge or recharge with a power loss ratio of λ. Then, the power discharge and recharge can be derived by $S_k^{input}(t) = (1 -$

$\lambda) \sum_{i \in G} [G_{ik}(t) - \Delta G_{kj}(t)]$, and $S_k^{output}(t) = \sum_{j \in D} G_{kj}(t)/(1-\lambda)$, respectively. The status of storage device can be derived by,

$$S_k(t) = \int_0^t (S_k^{input}(t) - S_k^{output}(t))dx, \tag{4.3.5}$$

where $0 \leq S_k(t) \leq S_k^{max}$ and S_k^{max} is maximum capacity of storage device k. When the storage device is integrated into the power grid, to achieve the minimization of power cumulative goal, the power distribution can be derived by Equation (4.3.4), which is represented by,

$$\textbf{\textit{Objective.}} \quad Min\left\{C_i \cdot G_T(t) + C_p \cdot \Delta G_T(t)\right\} \tag{4.3.6}$$

S.t.

$$\begin{cases} \forall i \in G \bigcup S, \forall j \in D \bigcup S, \forall k \in S, \\ G_{ij}(t) \leq L_{ij}, \\ G_i^{min} \leq G_i(t) \leq G_i^{max}, \\ G_i(t) = \sum\limits_{j \in D \bigcup S} G_{ij}(t), \\ G_T(t) = \sum\limits_{i \in G \bigcup S} G_i(t), \\ D_j(t) = \sum\limits_{i \in G \bigcup S} [G_{ij}(t) - \Delta G_{ij}(t)], \\ \Delta G_T(t) = \sum\limits_{i \in G \bigcup S} \sum\limits_{j \in D \bigcup S} \Delta G_{ij}(t) + \sum\limits_{k \in S} (\frac{\lambda}{1-\lambda} S_k^{input}(t) + \lambda S_k^{output}(t)). \end{cases}$$

Based on Equation (4.3.6), when $G_T(t) - \Delta G_T(t) > \sum\limits_{j \in D} D_j(t)$ (i.e., the total power supply to users is more than the total power demand), the storage device should recharge the power from the grid. When $G_T(t) - \Delta G_T(t) < \sum\limits_{j \in D} D_j(t)$ (i.e., the total power supply to users can not satisfy the total power demand), the storage device will discharge to the grid to make up the lack of power in the grid. Notice that,

in comparison with distributed energy resources without storage devices, the integrated energy storage power grid could make the power generation less than the user demand.

4.3.2 User Service Reliability

From the cumulative power cost analysis discussed in Section 4.3.1, we know that the integration of distributed energy resources could reduce the cumulative power cost. Therefore, it is critical to investigate the impact of the user service reliability when distributed energy resources are integrated into the grid. Here, we consider the user service reliability can be affected by two factors: (i) user demand, and (ii) power generation by distributed energy resources. In the following, we give Definitions 3 and 4 to define the effect by these two factors.

Definition 3: The Successful Demand Ratio is defined as the probability that with a certain power supply as threshold, users can get enough power to satisfy their demands in a time interval, which can be represented as $DR(t)$.

Definition 4: The Effective Generation Ratio is defined as the probability that a quantity of power supply can be provided by distributed energy resources in a time interval, represented as $RR(t)$.

Based on these two definitions, the user service reliability can be formalized as $SR(t) = DR(t) \cdot RR(t)$.

In the following, we will analyze the user service reliability in those three systems, respectively.

4.3.2.1 Traditional Bulk Power Generation

In our preliminary results, we conduct the statistical analysis using Nonparametric test [105] and Q-Q plot test [106] on real-world historical data and show that we could approximate the distribution of meter data with a Gaussian distribution. Hence, the user demand could be estimated by a Gaussian distribution $N[\mu_j(t), \sigma_j^2(t)]$ in a time interval. Based on the Gaussian distribution, the probability of the total user demand will be $N[\mu(t), \sigma^2(t)]$, where $\mu(t) = \sum_{j \in D} \mu_j(t)$ and $\sigma^2(t) = \sum_{j \in D} \sigma_j^2(t)$.

Recall that in Definition 3, we assume that the user demand cannot be less than 0 and the cumulative probability of user demand from 0 to ∞ will be 1. Assume that the total power supplied by the traditional power generator is M, if user demand is less than M, user can successfully get enough power. We assume the total user demand is $D(t)$. The successful demand ratio can be formulated as

$$DR(t) = \frac{\int_0^M e^{-\frac{[D(t)-\mu(t)]^2}{2\sigma^2(t)}} d[D(t)]}{\int_0^\infty e^{-\frac{[D(t)-\mu(t)]^2}{2\sigma^2(t)}} d[D(t)]}. \tag{4.3.7}$$

In a power system, without power generation by distributed energy resources, user service reliability can be regarded as the successful demand ratio, where $SR_1(t) = DR_1(t)$.

4.3.2.2 Integrated Distributed Power Generation

Recall that in Definition 4, we consider that power generation cannot be less than 0 and the cumulative probability of user demand from 0 to ∞ will be 1. We assume that the distributed power generation should be N in order to guarantee the grid demand. Then, when the distributed power generates more than N unit power, the grid demand will be satisfied. We assume that the total distributed power generation is $G(t)$. Hence, the effective generation ratio can be represented by

$$RR(t) = \frac{1 - \int_{-\infty}^{N} e^{-\frac{[G_T(t)-\mu'(t)]^2}{2\sigma'^2(t)}} d[G(t)]}{\int_{0}^{\infty} e^{-\frac{[G(t)-\mu'(t)]^2}{2\sigma'^2(t)}} d[G(t)]}. \tag{4.3.8}$$

Recall that in Definition 2, we assume that distributed power supply is N_1 and traditional power supply is N_2. Then, the user service reliability is

$$\begin{cases} SR_2(t) = DR_2(t) \cdot RR_2(t), \\ DR_2(t) = \frac{\int_{0}^{N_1+N_2} e^{-\frac{[D(t)-\mu(t)]^2}{2\sigma^2(t)}} d[D(t)]}{\int_{0}^{\infty} e^{-\frac{[D(t)-\mu(t)]^2}{2\sigma^2(t)}} d[D(t)]}, \\ RR_2(t) = \frac{1-\int_{-\infty}^{N_1} e^{-\frac{[G(t)-\mu'(t)]^2}{2\sigma'^2(t)}} d[G(t)]}{\int_{0}^{\infty} e^{-\frac{[G(t)-\mu'(t)]^2}{2\sigma'^2(t)}} d[G(t)]}, \end{cases} \tag{4.3.9}$$

Based on Equation (4.3.9), when $DR_1(t) < DR_2(t) \cdot RR_2(t)$, i.e., $\frac{DR_1(t)}{DR_2(t)} < RR_2(t)$, the user service reliability will increase. The increased number of distributed energy resources will increase the stable power generation $\mu'(t)$ and error variance $\sigma'^2(t)$, where $\mu'(t) < \mu''(t)$ and $\sigma'^2(t) < \sigma''^2(t)$. With the increase of $\mu'(t)$ and $\sigma'^2(t)$, we know that $\frac{N_1-\mu'(t)}{\sigma'^2(t)} < \frac{N_1-\mu''(t)}{\sigma''^2(t)}$ and $\phi[\frac{N_1-\mu'(t)}{\sigma'^2(t)}] < \phi[\frac{N_1-\mu''(t)}{\sigma''^2(t)}]$. Then, from Equation (4.3.8), the effective generation ratio will increase, leading to increase of user

service reliability.

4.3.2.3 Integrating Both Distributed Power Generation and Storage

Recall that in Section 4.2, standing as a buffer between power generators and end users, the energy storage will provide stability to the power grid. Based on Equation (4.3.9), when distributed energy resources and storages are integrated, the user service reliability can be represented as,

$$\begin{cases} SR_3(t) = DR_3(t) \cdot RR_3(t), \\ DR_3(t) = \dfrac{\int_0^{N_1+N_2+S} e^{-\frac{[D(t)-\mu(t)]^2}{2\sigma^2(t)}} d[D(t)]}{\int_0^{\infty} e^{-\frac{[D(t)-\mu(t)]^2}{2\sigma^2(t)}} d[D(t)]}, \\ RR_3(t) = \dfrac{[1-\int_{-\infty}^{N_1} e^{-\frac{[G(t)-\mu'(t)]^2}{2\sigma'^2(t)}} d[G(t)]]}{\int_0^{\infty} e^{-\frac{[G(t)-\mu'(t)]^2}{2\sigma'^2(t)}} d[G(t)]}, \end{cases} \tag{4.3.10}$$

From Equations (4.3.9) and (4.3.10), we can observe that integrating storage devices do not affect the effective generation ratio. The total power supply capacity will increase due to the storage device discharging power to the grid, where $N_1 + N_2 < N_1 + N_2 + S$. As a result, the successful demand ratio will increase, which will cause the user service reliability to increase as well. From another point of view, the storage could balance the total power supply in the grid and smooth the bulk power generated in different time intervals. This makes up the impact of distributed energy resources on the user service reliability.

4.4 Performance Evaluation

In this section, we show the evaluation results of the three power grid systems in terms of cumulative power cost and user service reliability discussed in Sections 4.2 and 4.3.

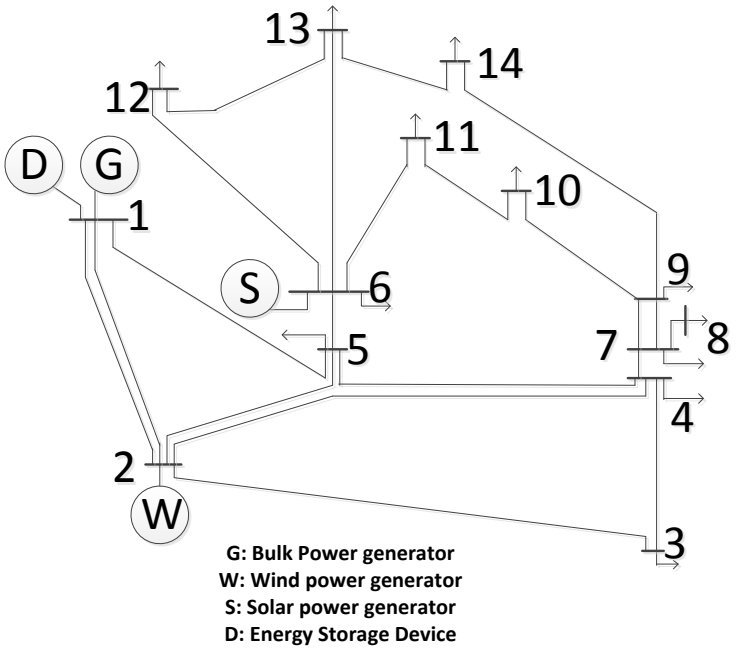


Figure 4.1: IEEE 14 Bus Topology

To demonstrate the effectiveness of integrating distributed energy resources into the power grid, we conduct the performance evaluation based on the IEEE14-Bus topology as shown in Figure 4.1. In this topology, we deploy one bulk power generator and storage device at bus 1, wind power generator at bus 2, and solar power generator at bus 6. The connections between these buses are the power transmission lines. Notice that such a simplified topology can validate our analysis and show

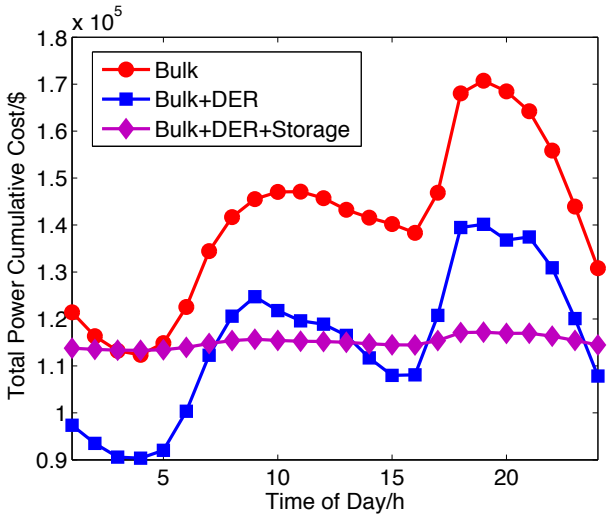


Figure 4.2: Total Power Cumulative Cost of Three Systems

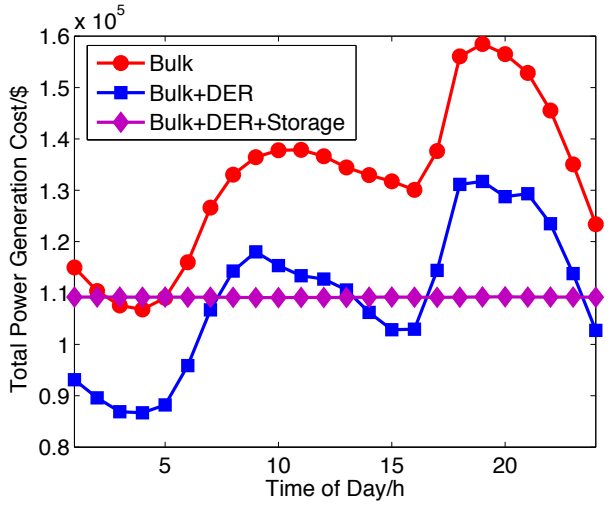


Figure 4.3: Total Power Generation Cost of Three Systems

the consequences.

In our simulation, we use the data set from California ISO website [101]. We consider the power line voltage is $400KV$. We consider the copper power transmission, whose resistivity is $0.000999ohms$ and cross section is 2.081 square millimeters [107]. We select $500MW$ capacity copper hexacyanoferrate battery as an energy storage device, in which power loss ratio is only 1% in energy transform process [108]. We consider the marginal generation cost of the bulk generator is $50\$/MWh$, and the marginal generation cost of wind and solar generators is $1\$/MWh$, where the marginal generation cost is defined as the cost of $1MWh$ per unit power generation. We assume that the power loss marginal cost is average energy price of the day, which is $100\$/MWh$ [109].

Recall that we consider the following metrics discussed in Section 4.3 using in the Matlab simulation: (i) *Cumulative Power Cost:* It is defined as the total cost of the power generation and distribution process in a time interval. (ii) *User Service Reliability*: It is defined as the probability of user energy demand to be met by the grid.

Figure 4.2, Figure 4.3, and Figure 4.4 show the change trend of the cumulative energy cost, power generation cost, and power distribution cost of three systems we introduce in Section 4.2 in 24 hours. As we can see, with the different time of the day, the trend of cumulative energy cost is fluctuant. Obviously, the cumulative

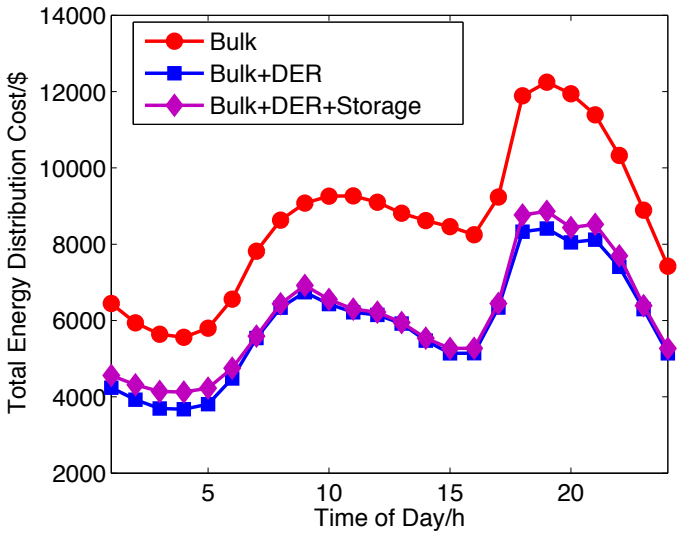


Figure 4.4: Power Transmission Loss of Three Systems

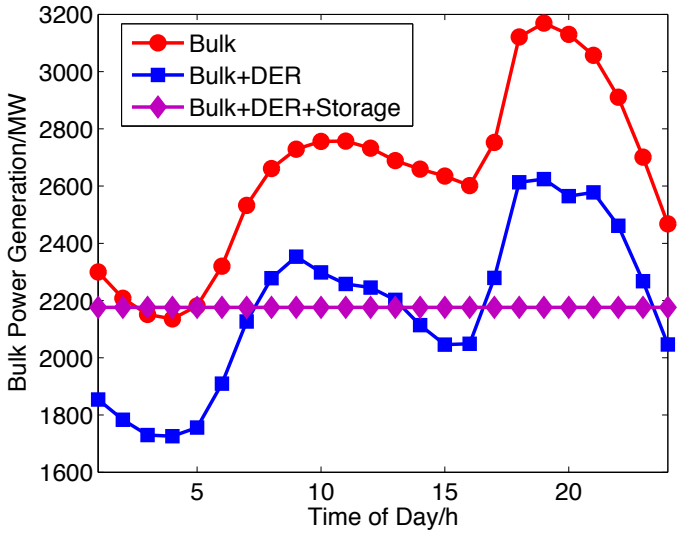


Figure 4.5: Comparison of Power Generation from Bulk Power Generator

energy cost of the power grid with integrated distributed energy resources is about 4×10^5\$ more than that of the bulk power grid, which is because the marginal power generation cost of the distributed energy resources is low and distributed energy resources that are nearby users can reduce the power distribution cost. When the 500 MW capacity storage is integrated into the grid, the cumulative energy cost curve becomes stable at 1.15×10^5\$, indicating that the storage device could reduce fluctuate of the cumulative energy cost and migrate the cumulative energy cost in peak time to that of no-peak time.

Figure 4.5 illustrates power generation from bulk power generator of three systems in the time of day. As we can see, the integration of distributed energy resources can help the reduction of bulk power generation and the energy storage will smooth the trend of bulk power generation that bulk power generation can be maintained at 2200 MW. The result matches with the analytical result in Section 4.3.

Figure 4.6, Figure 4.7, and Figure 4.8 show the user service reliability of three systems. From Figure 4.6, we observe that, when the bulk power generation approaches 3000 MW, user service reliability approaches 100%. Comparing with Figure 4.6, when the bulk power generation is 2000 MW and the distributed power generation is 1000 MW, the user service reliability in Figure 4.7 is higher than that

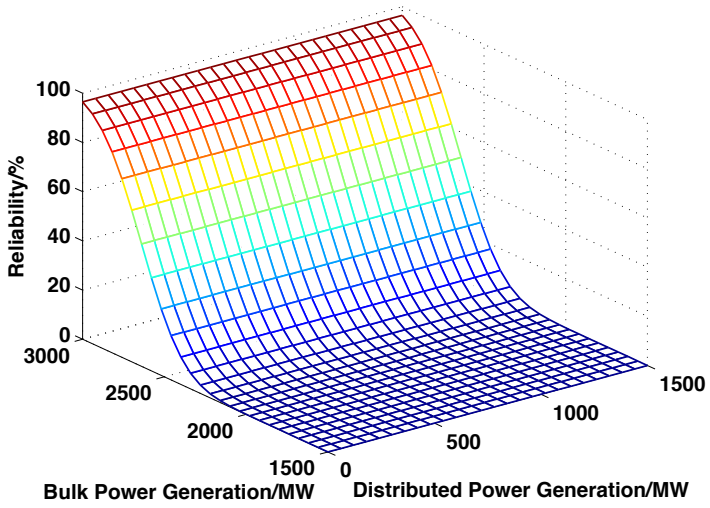


Figure 4.6: User Service Reliability of Traditional Bulk System

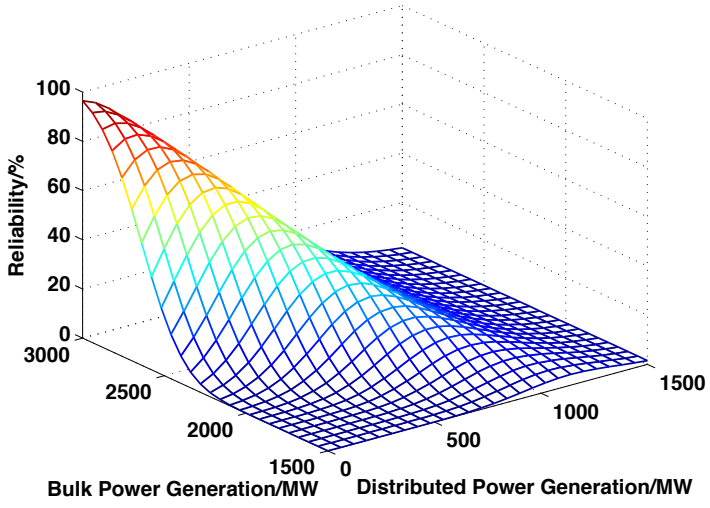


Figure 4.7: User Service Reliability of Integrated DER

of only 2000 MW bulk power generation in Figure 4.6 because of the increase of total power generation. In comparison with Figure 4.7, by maintaining the distributed power generation at 500 MW, and bulk power generation at 2000 MW, Figure 4.8 shows that integrating energy storage could achieve a higher user service reliability.

Figure 4.9 depicts tradeoffs between the user service reliability and the number of distributed energy resources. As we can see, when we assume that the amount of the distributed power generation is constant, the more distributed energy resources are integrated to the grid, the effective generation ratio will increase, leading to the increase of user service reliability.

In addition, in Figure 4.10, we consider the optimal storage device deployment. With different storage capacities (e.g., battery based on Li-ion technology (50 MW), Adiabatic compressed-air energy storage-CAES (100 MW), and Pumped hydroelectric storage-PHS (300 MW)) and serviceability options required, the deployment decision and the storage device deployment locations are different. The numbers 1 and 0 on the z axis represent the deployment decision: either deployment or no deployment, respectively. When the capacity is 50 MW, to achieve a high serviceability, more storage devices should be deployed in the power grid. For example, to achieve 98% serviceability with 50 MW storage devices, each bus should be deployed with a storage device. When the storage capacity is 300 MW, which can achieve high serviceability, the minimal number of storage devices is 4 and the storage devices

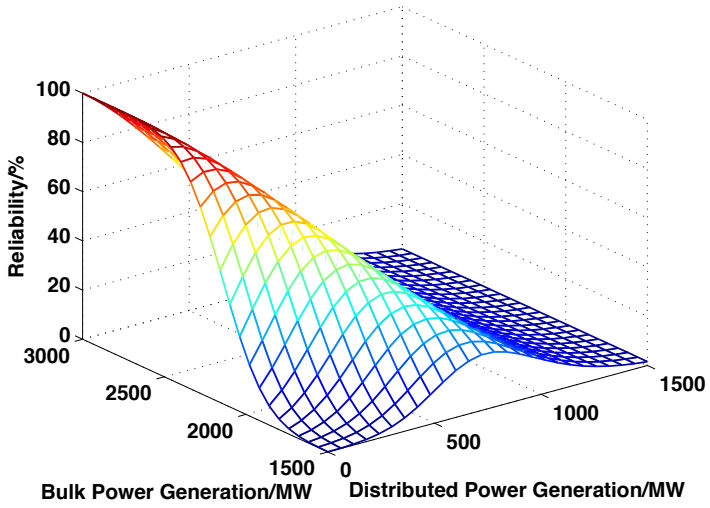


Figure 4.8: User Service Reliability of Integrated Storage Devices

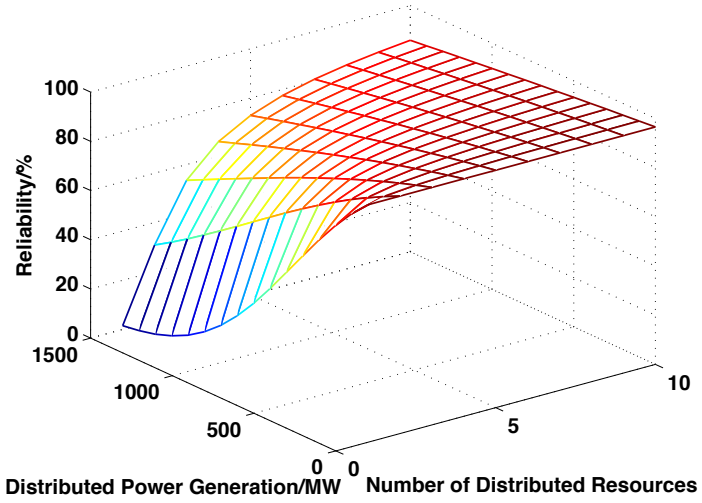


Figure 4.9: User Service Reliability vs. The Number of DER

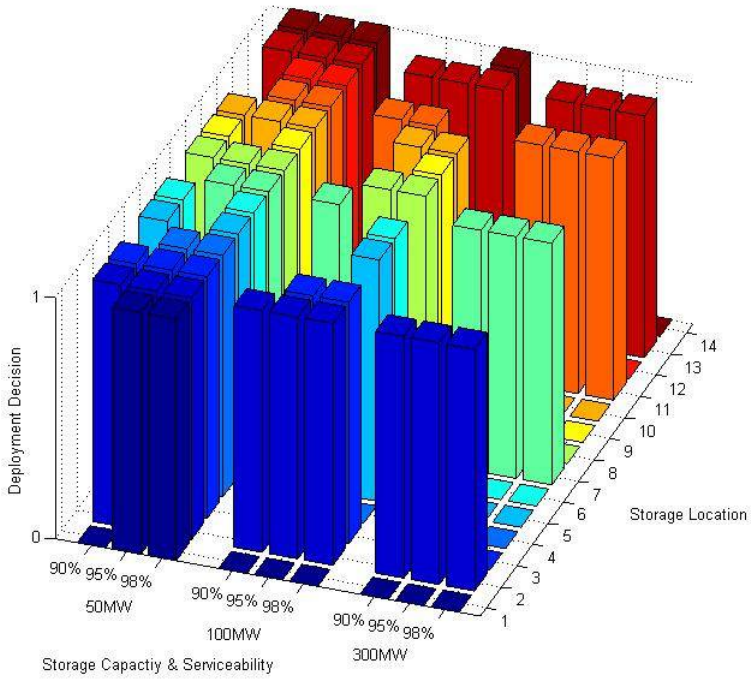


Figure 4.10: Storage Devices Deployment Location

should be deployed in buses 2, 7, 11, and 13, respectively.

4.5 Summary

In this chapter, we investigated the impact of integrating distributed energy resources and energy storage devices on the bulk power generation in the Smart Grid. We first introduced three types of power grid systems: traditional bulk power grid, power grid with distributed energy resources, and power grid with both distributed

energy resources and storage devices. We then defined metrics and formally analyzed the effectiveness of integrating distributed energy resources and energy storage devices into the bulk power generation. We conducted extensive simulations. Our simulation data show that integrating distributed energy resources conjointly with energy storage devices could reduce the generation costs, smooth the curve of bulk power generation over time, reduce bulk power generation, and distribution losses, and provide sustainable user service reliability.

Chapter 5

Secured Energy Resource Management in The Smart Grid

5.1 Overview

The Smart Grid [3] is expected to achieve a more efficient, reliable, secure, and resilient system operation and provide better service to customers than traditional power grids, by leveraging advanced cyber computer and communication technologies [110]. On the transmission and distribution side, Supervisory Control and Data Acquisition (SCADA) systems collect real-time information that provides wide area situational awareness of power grid status. On the user side, more precise real-time estimates of anticipated usage through Advanced Metering Infrastructure (AMI) enable demand response controls that could increase the efficiency of energy use.

While major research efforts have been conducted in improving the operational efficiency and reliability of power grids through the use of cyberspace computing and communication technologies, the risks of cyber breach on power grid systems need to be seriously investigated before the massive deployment of the Smart Grid technologies. There are growing concerns in the Smart Grid on protection against the malicious cyber threats [5, 111, 112]. The operation and control of Smart Grid largely depend on a complex cyberspace of computers, software, and communication technologies. An adversary has the potential to pose great damage to the grid,

including extended power outages, destruction of electrical equipment, increased energy cost and price, etc., if the system is compromised. Because the measurement component supported by smart equipment (e.g., smart meters and sensors) plays an important role, it can be a target for attacks. Notice that those measuring devices may be connected through the open network interfaces and the lack of tamper-resistance hardware increases the possibility of being compromised by the adversary [113, 114]. The adversary may modify data and compromise measuring components through injecting malicious codes into the memory of measuring components [115].

Nonetheless, developing secured energy resource management in the Smart Grid is challenging because of the following reasons: First, the Smart Grid is a highly distributed system, which inherently operates under the presence of various uncertainties posed by different types of failures and malicious attacks. Attacks can be from either cyber or physical grid components. It is challenging to quantify the impact of uncertainties from failures and threats and develop mechanisms to deal with those uncertainties. Second, Smart Grid is a very large and complicated system, which consists of many different functional components [3]. Therefore, how to investigate systematically the impact of attacks on the energy resource management and how to develop effective countermeasures are challenging problems.

To address those problems, in the chapter, we made several contributions. First,

we systematically explore the space of attacks in the energy management process, including modules being attacks (e.g., end-user, communication network, system operations, etc.), attack avenue (e.g., confidentiality, integrity, availability, etc.), attack strength (e.g., strong, stealthy) and system knowledge (e.g., full, partial). Specifically, we take the attacks against distributed energy transmission as a case study to investigate the risk of those attacks. Second, we develop the defense taxonomy to secure energy management with three orthogonal dimensions: methodology, sources, and domains.

5.2 Threat Taxonomy in the Smart Grid

Monitoring and controlling physical power system through geographically distributed sensors and actuators have become an important task in the Smart Grid. Nonetheless, the Smart Grid may operate in hostile environments and the sensor nodes lacking tamper-resistance hardware increases the possibility of being compromised by the adversary [61, 116]. Therefore, the adversary can launch attacks to disrupt the operation of the Smart Grid through the compromised meters and sensors.

While most existing techniques for protecting the Smart Grid have been designed to ensure system reliability (e.g., against random failures, etc.), recently there has been growing concerns in the Smart Grid on the protection against malicious cyber attacks [5, 111, 112]. It was found that an adversary may launch

attacks by compromising meters, hacking communication networks between meters and SCADA systems, and/or breaking into the SCADA system through a control center office LAN [117]. The adversary can inject false measurement reports to the controller through the compromised nodes. Therefore, the adversary can inject false measurement reports to disrupt the Smart Grid operations through compromised meters and sensors. Those false data injection attacks lead to the dangerous threats to the system. For example, one of the most important reasons for the 2003 Eastern blackout is that the state estimation programs for key areas were abnormal and failed to provide the system operators with the correct state information [118]. Stuxnet malware found in July 2010 targeted SCADA system in the process control system raises new questions about power grid security [119]. In addition, Liu *et al.* [5] investigated the false data injection attack, which can bypass the existing bad data detection schemes and arbitrarily manipulate the states of grid system, causing the control center to make wrong decisions on operating the power-grid network. Nonetheless, the risk and impact of attacks against energy management in the Smart Grid have not been studied in the past.

To address this issue, we propose a theoretical framework to explore the space of attacks, which is illustrated in Figure 5.1. Here, X axis represents modules being attacks $<X_1$: end user, X_2 : communication network, X_3 : system operation$>$. Y axis represents attack avenue $<Y_1$: confidentiality, Y_2 : integrity, Y_3 : availability$>$. Z axis

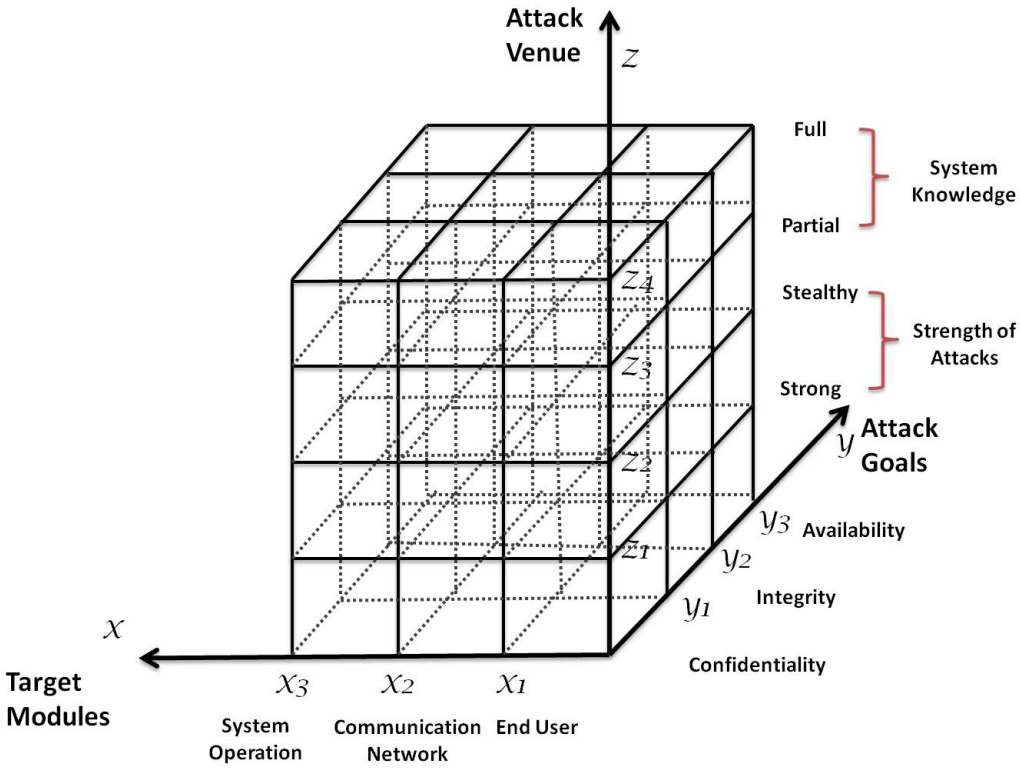


Figure 5.1: 3D Attack Space

represents attack strength - either <Z_1 : strong, Z_2 : stealthy> or <Z_3 : full system knowledge, Z_4 : partial system knowledge>. In the following, we introduce this framework in detail.

5.2.1 Target Modules

In dimension X, attacks targeting different functional modules could be systematically investigated, including end user, communication network, and system operation.

- *End User:* AMI is to provide the Smart Grid utilities with real-time energy data

and enable users to access real-time energy pricing and consequently make informed decisions about the user energy usage. AMI services include real-time energy measurement, power outage notification, power quality monitoring, and automated meter reading. AMI is a major component of the modern electric grid, which requires strong security features to prevent electric grid disruption. Major security issues are posed by the deployment of smart meters [120], including: (i) energy fraud where meter reading can be manipulated by returning false reading from credit meters or by forging the data in smart meters leading to economic losses, (ii) privacy attacks where sensitive and personally identifiable information in smart meter data can be hijacked or disclosed to unauthorized people to derive habits and behavior of the consumer.

- *Communication Network:* To achieve its intended goals, the Smart Grid requires a robust, resilient, and efficient communication network. Attacks on communication networks in the Smart Grid can be vulnerability-driven attacks that deal with the malfunction of network device or communication channel, data injection attacks which target the integrity of data exchanged in the network or intentional attacks where the adversary has a full understanding of the network topology. Using the time synchronization as an example, time synchronization is a core component of reliable automation, fault analysis and

recording in the Smart Grid and other critical infrastructure applications. One possible attack is to break the assumption that the link delay between master and slave nodes in IEEE1588 is symmetric, in which the random delays in Sync message from master node and DelayReq message from slave node will be added [121].

- *System Operation:* SCADA systems accomplish the management with the real-time access of local or remote data and channels transmitting the data to control center. SCADA systems are composed of computers, remote station control devices used for data acquisition. Nowadays, SCADA highly rely on open connectivity to corporate network and Internet for advantages in real-time efficiency and productivity. Therefore, the computerized real-time processes are subject to malicious attack while SCADA systems take little measures against these cyber threats. The attacks on SCADA systems have been widely studied in recent researches [55, 122–125]. The reported attacks are basically categorized based on several security properties such as timeliness, availability, integrity, and confidentiality. In particular, the attacks on the implementation of protocols target on the vulnerabilities in common protocols in SCADA (e.g., MODBUS, DNP3). One example is Stuxnet, the first publicly-known malware to exploit vulnerabilities in SCADA systems [59, 126].

5.2.2 Attack Goals

In the Y dimension, the attack goals in the Smart Grid consists of confidentiality, integrity, and availability.

- *Confidentiality:* Confidentiality refers to a set of security rules and measures that limit and control access to information. To be more explicit, confidentiality aims to protect personal privacy and proprietary information by preventing unauthorized disclosure of information while ensuring that data are accessible only to authorized personnel. In a scenario of attacks against confidentiality, the adversary will eavesdrop on communication channel in order to obtain the needed information. Previously, Li *et al.* [127] leveraged the information theory to study the communication capacity under an eavesdropping in the Smart Grid. Based on the information theory, Sankar *et al.* [128] developed the concept of competitive privacy and modeled privacy issues in the Smart Grid.

- *Integrity:* Integrity refers to a set of security measures that prevent unauthorized modification of data and information. The purpose of integrity is to provide assurance that information is accurate, consistent, and trustworthy over its entire life cycle. The false data injection attack is one example of attacks targeting the integrity of the system. In this scenario, the adversary forges or manipulates the data to corrupt critical information exchange

and impair decision making processes in the Smart Grid. For example, Giani *et al.* [129] studied the data integrity of attacks and potential defense techniques against those attacks. Pasqualetti *et al.* [130] developed a distributed method for state estimation to counterattack the new trend of false data injection attacks. Vikovic *et al.* [131] proposed a network-layer protection schemes against stealth attacks.

- *Availability:* Availability is the guarantee that data and information will be accessible for control and use in a timely and reliable manner. The most common attack against availability is the denial-of-service (DoS) attacks where the adversary attempts to disrupt, delay, or corrupt the communication network to significantly degrade its performance. DoS attacks can be launched at different layers of the communication network, including physical layer, MAC layer, network layer, and application layer. For example, at the physical layer, the common DoS attack is the channel jamming attacks. Various research works proposed the countermeasures against channel jamming. For example, Liu *et al.* [132] have proposed a randomized differential Direct Sequence Spread Spectrum (DSSS) for Jamming-resistant wireless broadcast communication. Jin *et al.* [133] have demonstrated the vulnerability of SCADA systems to DoS attacks by evaluating the impact of event buffer flooding attack in Distributed Network Protocol (DPN3) controlled SCADA.

5.2.3 Attack Venues

Combined with attack dimensions X and Y, attack dimension Z is considered as well, including stealthy/strong attacks and attacks with full system/semi/zero system knowledge. The goal of strong attacks is to inflict maximum damage in the shortest possible time. A stealthy adversary, in contrast, is more interested in compromising the Smart Grid and manipulating data over a long period of time while avoiding detection. For example, for an attack with a behavior feature, we can obtain the measure to quantify the behavior, including entropy, statistical mean, standard derivation, and others. Nonetheless, to avoid being detected by statistical anomaly detection and other schemes, the adversary may become stealthy and only marginally change the attack behavior. For example, Dan *et al.* [134] studied stealthy false data attacks against state estimators in power systems. Esmalifalak *et al.* [135] investigated stealth false data attacks in the Smart Grid without prior knowledge of the topology. Prior system knowledge is important to the attackers. For example, with the complete or partial infrastructure information, the attackers can investigate the optimal strategy to select attack nodes or links in both cyber network and physical power grid to achieve the maximum damage.

5.3 Case Study: Attacks against Distributed Energy Transmission

From the point of view of demand response, power systems are made of four essential components: the supply nodes that provide energy, the demand nodes representing energy consumers, the bidirectional communication network for data and information exchange, and the power transmission link lines. Balancing the supply and demand can be formalized as an optimization problem. An efficient and comprehensive power balance optimization integrates various constraints. The first constraint is that the total power input of demand nodes shall be equal to the user demanded power. Besides that, the total power output of supply nodes cannot exceed its capacity. In addition, the power flows should not exceed the capacity and physical characteristics of energy links or power lines.

The false data injection attacks as described in our previous research studies [61] and shown in Figure 5.2, can be classified as follows: $<X_1$: end user/X_2 : communication network/X_3 : system operation, Y_2 : integrity, Z_2 : strong/Z_4 : full system knowledge$>$. In a false data injection attack scenario, the adversary may alter the integrity of the demand and response message by forging energy information or energy link state which will cause a disruption of the energy distribution process and impair the optimization of power flows. There are the following three different ways to inject false data:

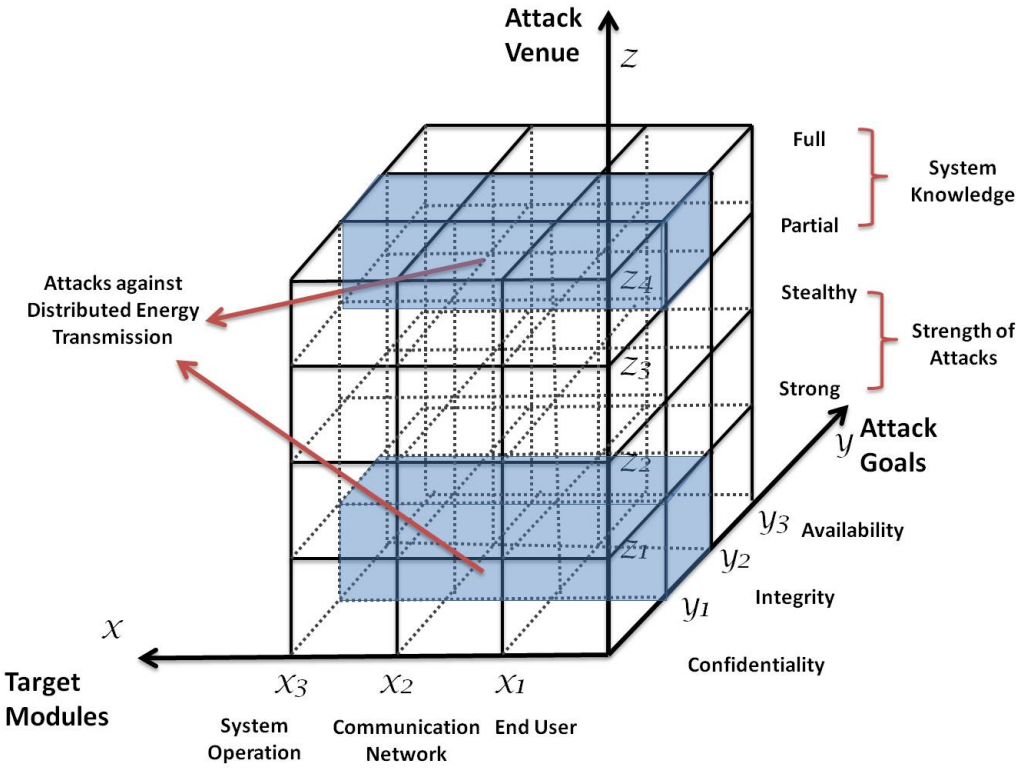


Figure 5.2: Attacks against Energy Transmission in 3D Attack Space

- *Injecting false energy request:* Once the adversary has compromised the demand nodes, a forged energy request that is more than its actual demand will be sent out. The resulting redundant power generation will lead to more energy waste partly because the excess power outweighs the limited capacity of the storage devices. In addition, the redundant power generation will increase the transmission cost and probably cause power outage as the power flows are distributed to the compromised demand nodes.

- *Injecting false energy demand:* In this case, the adversary uses the compromised nodes to forge less power supply than its actual capacity. The consequence will be disrupted power distribution and impaired power flows where some demand nodes fail to receive the enough power.

- *Injecting false link-state:* In addition to forging the energy information, the adversary can launch attacks by injecting false energy link-state information into the energy distribution. The consequence could incur high energy transmission cost and imbalance of energy supply because power flows will be routed to invalid energy links. In addition, as some energy links are falsely invalid, multiple nodes are isolated from the grid.

The graph and optimization theory can be leveraged to model the aforementioned attacks and to perform the quantitative analysis of their impact on the energy distribution. To demonstrate the impact of energy deceiving attacks on the distributed energy routing, we conducted performance evaluation based on the simplified version of the US Smart Grid shown in Figure 5.3. We select one major city of individual states as a node in the topology. The backbone of the interstate power delivery is based on the connection between these nodes. The fifty US states are selected as simulation objects, which are divided into five regions during simulations as shown in Figure 5.3. Notice that such a simplified topology could validate our ideas and demonstrate the consequences of those attacks.

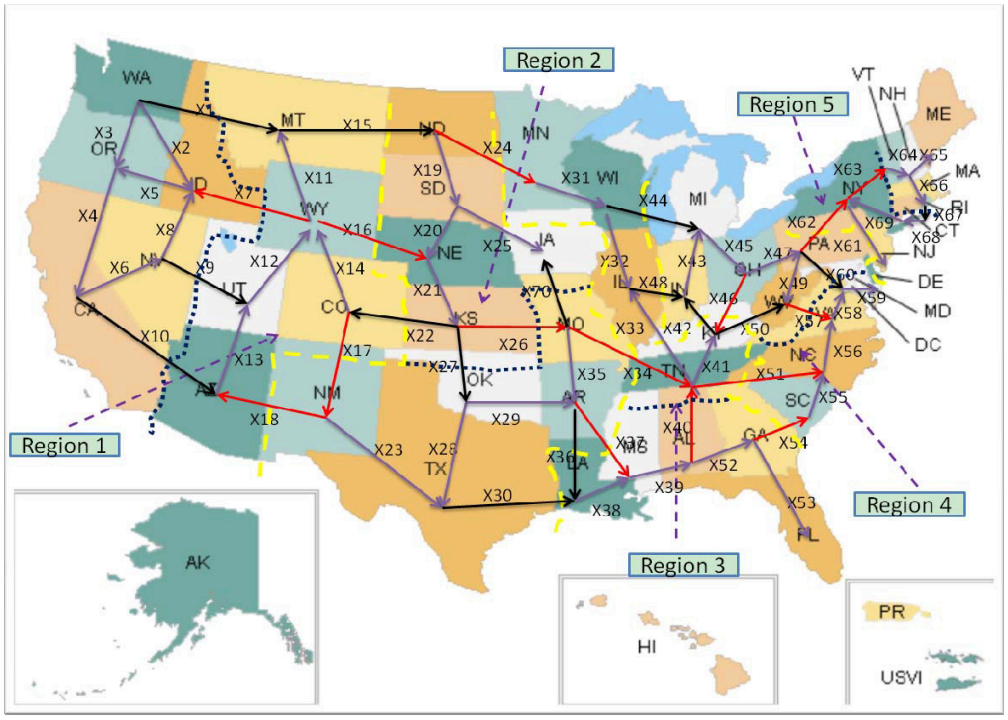


Figure 5.3: The US Smart Grid Topology

The data set used for simulations is based on “2009 US Energy Information Administration State Electricity Profiles" [136]. To access the capacity of power link, the averaged real-time data per second on each link, is computed based on the averaged 2009 US interstate power delivery data. The length of the power link, which represents the distance between two paired nodes, is computed using the *Google* map. To evaluate the impact of energy deceiving attacks against the distributed energy routing, we consider the following attacks discussed in previous: (i) the false quantity of power that demand-nodes provide, (ii) the false quantity of power that supply-nodes provide, and (iii) the false states of the power links. In

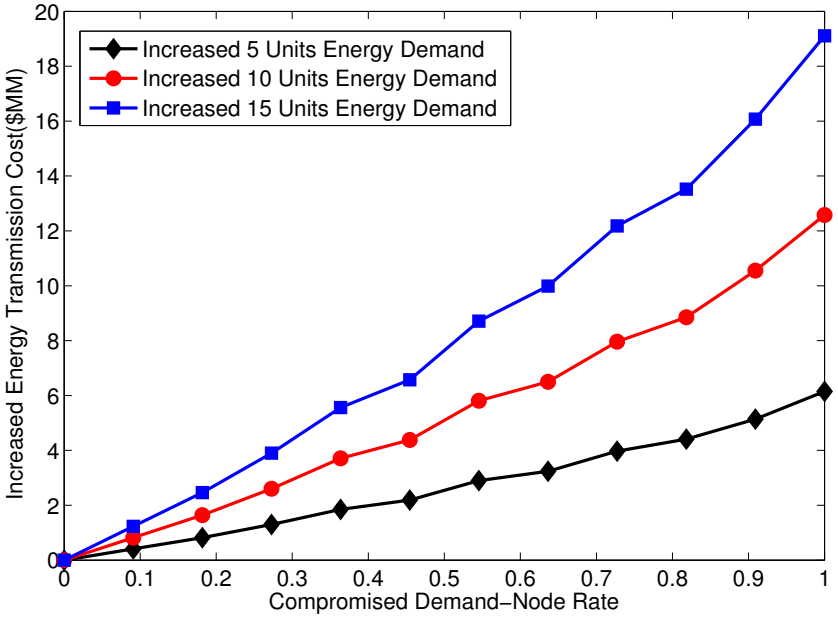


Figure 5.4: Increased Cost versus Compromised Demand-Node Rate

addition to the number of compromised nodes, another parameter to measure the strength of attacks is the quantity of power data to be manipulated. Obviously, the larger the quantity of power data to be manipulated, the stronger the attack is.

To measure the impact of energy deceiving attacks, we consider the increased delivery cost, the user outage rate, and supplied power loss as key metrics listed as follows : (i) *Increased delivery cost:* It is defined as the increased total power delivery cost caused by forged power data. (ii) *User outage rate*: With the manipulated quantity of power information, the total power supply may not satisfy all requests from nodes, and some nodes will become outage to ensure reliable power support of other nodes in the grid. (iii) *Supplied energy loss*: The forged power requests will

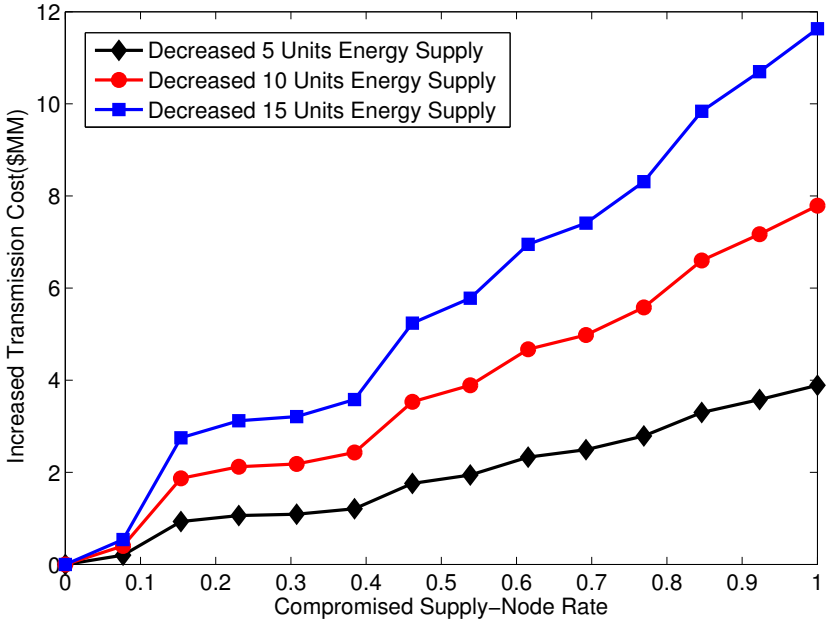


Figure 5.5: Increased Cost versus Compromised Supply-Node Rate

make the waste of the power supply. Our simulation data focus on these metrics to evaluate the impact of attacks. All simulations in this paper were conducted using Matlab 7.0.

Impact on Increased Power Delivery Cost: In this set of simulations, we designate 22 states of the US as power-demand states. Figure 5.4 shows the impact of the compromised demand-node rate on the increased power delivery cost. As we can see, with the increase in the compromised demand-node rate, the power delivery cost increases almost linearly. The different curves in the figure show the different quantity of forged power request. Obviously, the larger the quantity of power request data to be manipulated, the more power delivery cost is increased.

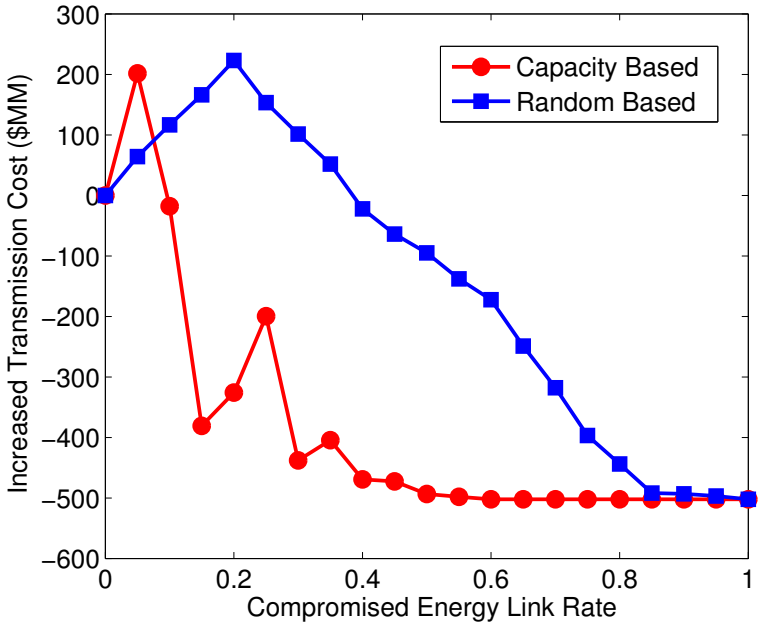


Figure 5.6: Energy Delivery Cost versus Compromised Energy Link Rate

When all demand nodes are compromised and forged 15 units power request, there is 19.11MM cost increase. The result matches with the our analytical result.

When the power-supply nodes are compromised, Figure 5.5 depicts the impact of the compromised supply node rate on the increased power delivery cost. As we can see, with the increase in the compromised supply-node rate, the power delivery cost increases as well. When more quantity of power is manipulated to reduce power supply, the power delivery cost becomes higher. For example, if 84.6% of supply-nodes are compromised, and each supply-nodes is manipulated to reduce 10 units supply, the increased power delivery cost approaches 6.6MM.

For attack forging the false state of power links, Figure 5.6 shows that when the

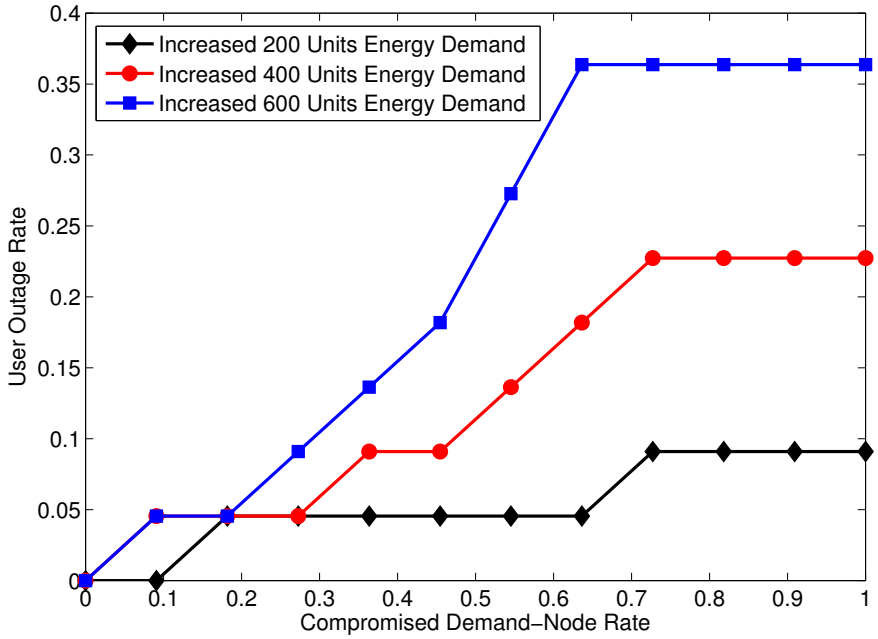


Figure 5.7: User Outage Ratio versus Compromised Demand-Node Rate

link is claimed as invalid, the attack impact on the power delivery cost. We can observe that, when the compromised power link rate is small, the power delivery cost grows linearly. With the increase in the compromised power link rate, the power delivery cost declines because of the declined number of power-demand nodes and power-supply nodes. For example, Figure 5.6 shows that the increased delivery cost would drop off, when 20% links are compromised. The data “20%" is obtained through the performance evaluation and it is related to the network setting (e.g., topology and others). When all states of links are manipulated as invalid, there is no energy delivery in the grid, so the energy delivery cost approaches zero. As shown in Figure 5.6, selecting compromised power links based on capacity leads

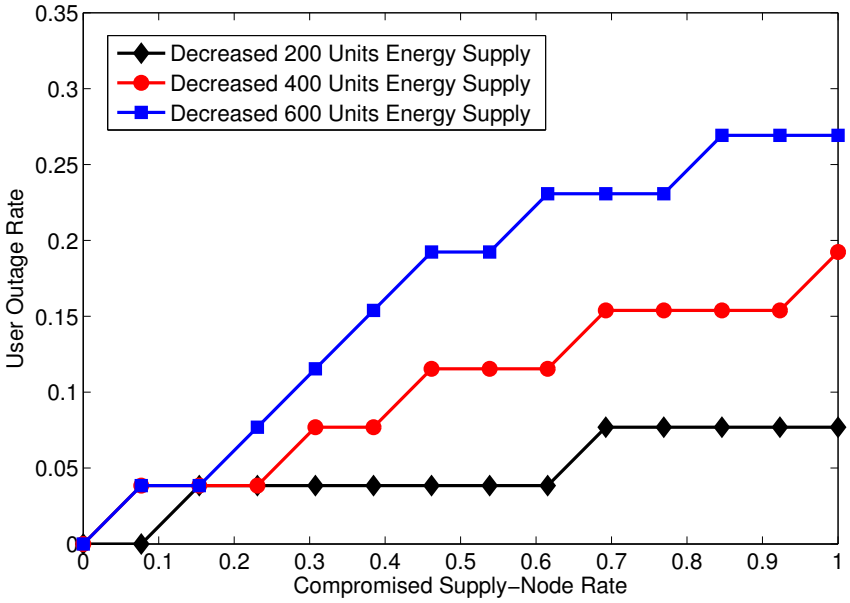


Figure 5.8: User Outage Rate versus Compromised Supply-Node Rate

to more serious effect on the delivery cost than selecting compromised links randomly, because randomly selecting compromised links will not always cause energy routing changes.

Impact on User Outage Rate: The forged power data will result in imbalance between power-supply and power-demand nodes. Figure 5.7 depicts the user outage rate versus the compromised demand-node rate. As we can see, when the quantity of power in compromised node is small, the user outage rate is not significant. When the large power requests caused by the attack, the user outage rate increases rapidly with the increase in the compromised demand-node rate. When

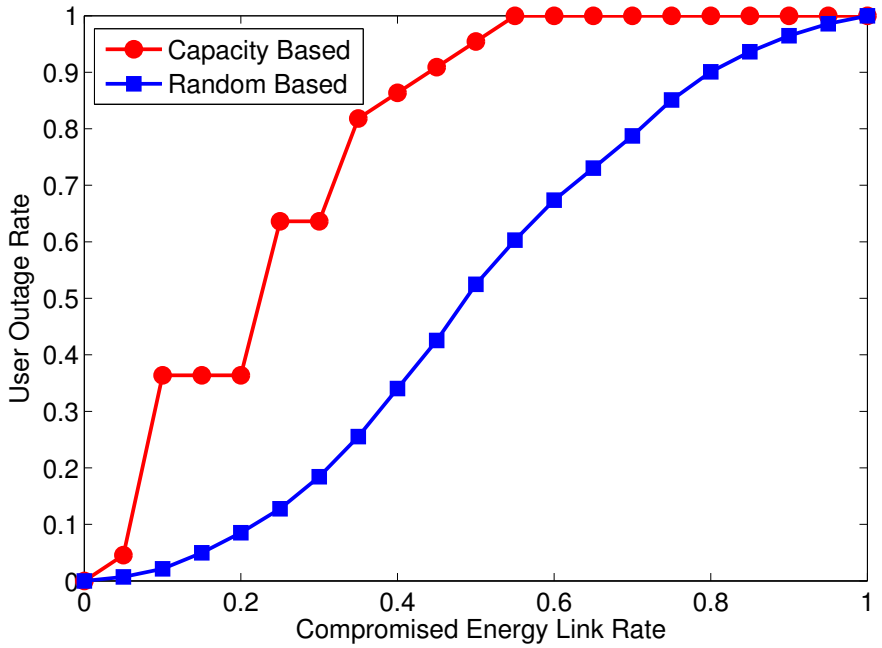


Figure 5.9: User Outage Rate versus Compromised Energy Link Rate

all demand-nodes are compromised, 600 units demand is manipulated at each compromised node, the user outage rate approaches 36.3%.

Figure 5.8 depicts the impact of user outage rate versus the number of supply-nodes being compromised. When a small number of supply-nodes are compromised or the small quantity of power supply is manipulated, the total power supply could meet the total power demand. Obviously, at the beginning of these curves shown in the figure, because there is no influence on demand-nodes, the user outage rate is almost zero. When around 15% supply-node is compromised, the user outage rate increases rapidly.

Figure 5.9 shows the user outage rate versus the compromised power link rate.

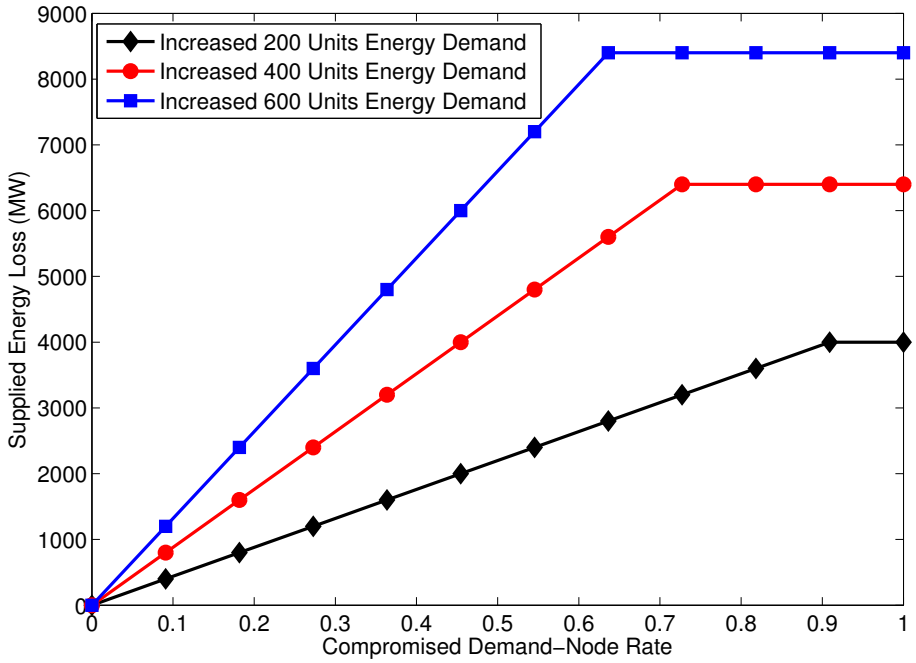


Figure 5.10: Supplied Energy Loss versus Compromised Demand-Node Rate

As we can see, with the increase in the compromised power link rate, the outage nodes increases smoothly at the beginning. This indicates that the small number of the power links claimed as invalid has a little impact on the effectiveness of energy routing process. Nonetheless, with the increase in the compromised power link rate, more nodes cannot obtain enough power from the grid, leading to more nodes to be outage. When the compromised link ratio is around 30%, the user outage rate grows rapidly. It also shows that selecting compromised power links based on capacity leads to greater user outage rate than selecting compromised links randomly.

Impact on Supplied power Loss: To investigate the impact of attacks on energy power loss, we show the relationship between the supplied power loss and the compromised demand-node rate in Figure 5.10. When the forged power request increases, more energy is supplied to meet these forged power requests. As we can see, the supplied power loss increases in a linear fashion with the compromised demand-node rate. When the supplied power loss approaches some level, it will stay in the same value. For example, for 600 units demand manipulated at each node, when all demand-nodes are compromised, the power supply loss will be $8400MW$, which is the same supplied power loss as the scenario, in which 63.6% of demand-nodes are compromised. Because the quantity of power requests is manipulated, more demand-nodes are outage. When compromised demand-nodes become outage, the forged power requests cannot be generated by these nodes.

5.4 Defensive Taxonomy

Based on the explored attack space in the Smart Grid, to defense against the diversified threats, as shown in Figure 5.11, we develop a defensive taxonomy in the Smart Grid, which consists of defensive methodology, defense sources, and defense domains.

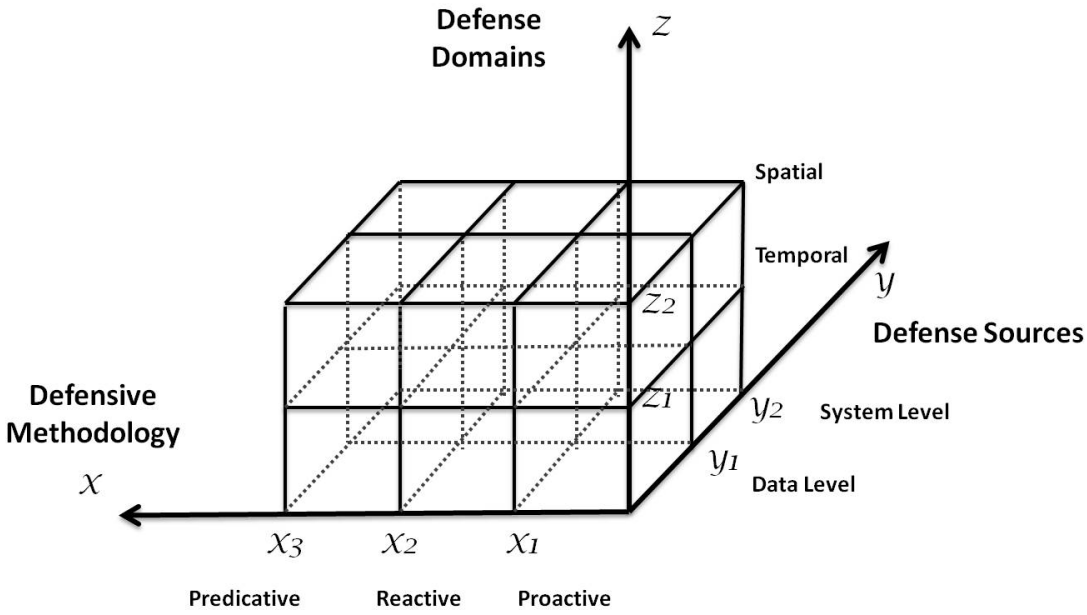


Figure 5.11: Defensive Taxonomy

5.4.1 Defensive Methodology

For the defensive methodology, we will introduce proactive defense, reactive defense, and predicative defense.

- *Proactive Defense:* The proactive defense is referred to as preemptive self-defense actions to interdict and disrupt attacks against the Smart Grid. Many proactive automatic defense technologies have been developed [137], in which emergent attack strategies can be anticipated and these insights are incorporated into the defense designs. The proactive defense takes preemptive self-defense actions to interdict and disrupt attacks against the Smart Grid. Numerous monitoring and detection tools (e.g., Fuzzing [138], SYSSTAT [139],

etc.) can be used to discover exploitable vulnerabilities proactively to make systems robust against cyber-attacks, in addition to anticipating potential causes of attacks. In addition, various system management and security tools can be deployed in the Smart Grid to establish a trustworthy architecture.

- *Reactive Defense:* Proactive defense mechanisms can mitigate and disrupt most of the known attacks against the Smart Grid. Nonetheless, the new cyber-attacks can still bypass proactive defense. To deal with new cyber-threats, reactive defense mechanisms, which consist of effective data attestation and anomaly detection, can be leveraged to diagnose the behaviors of Smart Grid systems based on the monitored metrics (e.g., state estimation etc.). For example, intrusion detection system can be deployed in the Smart Grid to conduct anomaly detection that monitors the characteristics of individual nodes and the events occurring in nodes for suspicious activities. The threat monitoring software or tools installed on nodes (e.g., smart meters, SCADA sensors) collect suspicious information in real-time from real-time data, system logs, security logs, application logs, and others, and forwards detection reports to the management node, which further conducts threat analysis and detection. The monitoring tools monitor suspicious activities on the node, including the integrity of system files, dynamic behavior, suspicious processes, illegal resource accesses, and others.

- *Predicative Defense:* Compared with reactive detection in which corrective actions are taken after an anomaly occurs and might prolong service downtime, predictive defense can achieve online anomaly prediction and raise advance anomaly alerts to system administers in a just-in-time fashion, aiming at raising advance alerts to trigger anomaly prevention. Generally speaking, predictive defense aims to improve the ability of defense systems to predict behaviors of new attacks. For example, effective techniques shall be developed to foresee impeding system anomalies through attacks forecasting.

5.4.2 Defense Sources

As we stated in Section 5.2, the adversary can launch various sophisticated attacks against Smart Gird to disrupt the system operations by injecting false data to the working nodes, distributing computational results for the Smart Grid applications. Therefore, to deal with different types of attacks, we propose our defense techniques in both data and system levels.

- *Data Level:* In the data level, security can ensure the trustworthy of sensing data such as meter readings. For example, we can develop a pre-defined data self-correction method to detect and recover the compromised computational data to achieve proactive defense. In addition, low-cost data attestation mechanisms can dynamically verify the integrity of data processing results and

pinpoint malicious nodes when inconsistent results are recognized.

- *System Level:* In the system level, real-time behaviors of system operations are examined to ensure that the system is in both operation secure and efficient. For example, as a proactive defense, we can implement and deploy monitoring and detection tools to discover exploitable vulnerabilities proactively to make smart gird system robust against cyber-attacks. For reactive defense, effective anomaly detection techniques need to be developed. One possible way is to use the behavior based detection approach through machine learning (e.g., Support Vector Machine, Naive Bayes, etc.) and other statistical based detection schemes (e.g., hypothesis sequential testing, etc.).

5.4.3 Defense Domains

Targeting the sensing data of the Smart Grid, the adversary can make large changes to the measurement data in a short period in territory or make small, subtle changes over a long time interval. To defense these attacks, we consider both spatial-based and temporal-based defense methods.

- *Spatial-based Defense:* Recall that in stealthy attacks, the adversary may change measurements from multiple sensors marginally, such that individual compromised measurements will not be detected by the statistical anomaly detection discussed above. Spatial-based detection scheme can be used to detect such

stealthy attacks. When we view all the measurements received at a certain time as a unity spatially, the accumulated deviation of all the marginally compromised measurements will be significant. The measurements are random which follows a multivariate Gaussian distribution and can be estimated based on historical data. The use of the Gaussian distribution can be theoretically justified by assuming that many small, independent effects are additively contributing to each observation. Nonetheless, when false data injection attacks exist, it must change some specific measurements marginally and the combination of those measurements will lead to the state variables derived far from their true values. The deviations of all the measurements can be accumulated in a vector from their means, and the accumulated deviation can be stood up. Considering the null hypothesis at a certain false positive rate, a threshold shall be given to detect the data changes.

- *Temporal-based Defense:* one of temporal based defense strategies is based on on-line nonparametric cumulative sum (CUSUM) change detection mechanism [54]. In this scheme, the two hypotheses will be defined: H_0 (normal condition) and H_1 (being attacked). The CUSUM change detection algorithm assumes the observation begins with H_0, and at time k, it changes to hypothesis H_1. The goal of this scheme is to detect such a change as soon as possible.

Given a suppressed false positive rate, the CUSUM algorithm tends to minimize the time N ($N \geq k$), for which the test stops and determines whether a change occurs. Two metrics are used to measure the effectiveness of the temporal-based detection: false positive rate and detection time. The false positive rate is defined as the probability of falsely rejecting the null hypothesis H_0 and the detection time is the average time that it takes to detect attacks. Obviously, the smaller the values of both metrics, the higher the performance of the detection.

5.5 Integrated Simulation and Emulation Environment

It is worth noting that our developed attack and defense taxonomies can establish a foundation to further explore possible attacks and defensive mechanisms. Using orthogonal to mathematical and logical methods, we can derive attack scenarios and defensive schemes based on our proposed taxonomies. To understand the consequence of attacks and study the effectiveness of countermeasures, we should develop integrated simulation (co-simulation) and emulation environment. Using the Smart Grid as an example, an integrated co-simulation platform should be developed to simulate and emulate both physical components (e.g., smart meters, inverters, sensors, relays, data acquisition devices, PMU (power management unit), etc.) and cyber components (e.g., SCADA, AMI, etc.). The high performance servers

can be deployed to execute a large number of virtual machines to build large-scale emulation testbed. Simulink, the virtualized software-based PMU, and network emulator (ns-3) can be integrated into the testbed. In each virtual machine, the specific software (e.g., powerworld, RTDS, Trilliant, Testbench, etc.) can be deployed to emulate energy-related information. Based on this integrated platform, a set of use case scenarios derived from our developed taxonomies can be evaluated. In this way, the attack impact on both power grid and cyber networks and the effectiveness of countermeasures to mitigate attack impact can be evaluated.

5.6 A Unified Theoretical Framework to Investigate the Effectiveness of the Synergy of Risk Analysis, Threat Detection, and Defense Reactions

Generally speaking, in the Smart Grid, the goal of risk analysis area research is to assess vulnerabilities and risks. The goal of a detection area is to conduct anticipating, detecting and analyzing malicious grid system activities. The goal of an agility area (as one of effective defense reaction) is to undertake agile cyber maneuvers to thwart and defeat malicious activities. With the synergy of these three areas, we can develop effective defense of the Smart Grid to increase the cost of launching successful attacks from the adversary and/or identifying attacks while achieving high network performance with sustainable maintenance cost. For example, the maneuvers in agility can be assisted to recognize the assets, threats and vulnerabilities

while conducting risk analysis and threat detection.

Nonetheless, there is no uniform framework to assess risks in the face of unknown system vulnerabilities. To this end, the control theoretical based framework shall be studied, enabling dynamical threat monitoring of system to achieve effective defensive actions. In addition, Markov chains and a game theoretical based modeling approach shall be investigated to systematically study the tradeoff between attack strategies and countermeasures.

5.6.1 Control Theory Based Theoretic Foundation

Control theory is an interdisciplinary system of research to study the behavior of dynamic systems having one or more inputs and outputs. A dynamic control system comprises an external input referred to as the reference and a controller that manipulates or compensates the inputs of the system to obtain the desired effect on the output. The aim of applying the control theory in the Smart Grid is to achieve the grid system stability through appropriate actions in the feedback loop with controller in response to disturbances and deviation from a set point. There are different types of control systems, including open loop control, closed loop control, feedback control, and feed-forward control systems. The open loop control has no feedback, and requires the input to return to zero before the output returns to zero. The closed loop control is a self-adjusting system and also has feedback. The feed-forward control is used to limit the deviation from the stability set point

and prevent disturbances. The feedback control is a reactive control that automatically compensates for disturbances and deviations. In the fast, there are numerous research efforts on applying control theory to network security [111, 140–142]. For example, Cramer *et al.* [140] described a concept in network intrusion detection based on the statistical recognition of an intruder's control-loop. Cardenas *et al.* [111] studied the problem of securing control and characterized the properties required by a secure control system, and the possible threats.

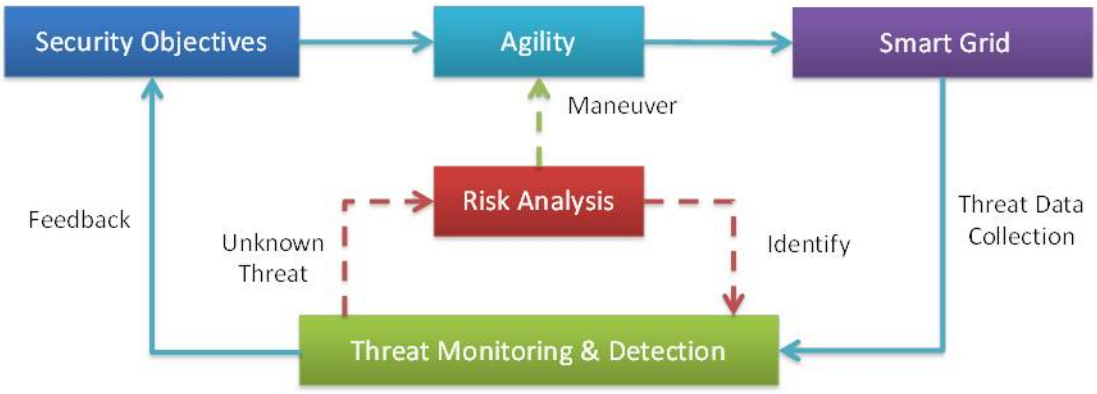


Figure 5.12: Control Model of Threat Monitoring and Detection Systeme

Different from the previous research efforts, we leverage the control theory to dynamically monitor and detect the security status of Smart Grid to make it achieve certain characteristics such as security controllability and observability in order to achieve rapid defense actions against attacks. As shown in Figure 5.12, the threat detection data can be collected through the monitoring agents deployed in the Smart Grid. To enable the fast feedback and threat detection in the feedback loop, a cloud based technique can be used to speed up the threat monitoring

and detection. To effectively detect threats with unknown signatures or features, dynamic risk assessment schemes are used to support threat identification and development of defensive mechanisms. Concurrently, comparing the detection results with predefined security objectives, the defense reaction (e.g., agility) of the system is to support the planning and control of a maneuver, such as intrusion detection deployment, firewall configuration, and filter configuration.

Through the control theory based framework, system stability from the security and system performance aspects can be studied. Specifically, principled theories leading to autonomous anticipation and adaptation to threats can be leveraged to eliminate costly, labor-intensive defensive measures and repairs to the Smart Grid. As a result, the complexity inherent to security can be significantly reduced and the impact on the Smart Grid operations can be controlled.

5.6.2 Markov Chains & Game Theoretic Based Approach

As we mentioned in Section 5.4, our proposed control theoretic framework can detect the deviation from the predefined system security objectives. Nonetheless, how to effectively adapt and defend strategies to force the adversaries to abandon their attacks or make those attacks less effective are challenging issues. To this end, we use the Markov chains and game theory to dynamically adapt detection and response strategies to deal with various cyber-attacks. In the past, game theory has been extensively studied in distributed systems and network security [143–

146]. For example, Alpcan and Basar [145] conducted a game-theoretic analysis of anomaly detection in access control systems, while Liu *et al.* [146] applied game theoretic results to anomaly detection in wireless networks.

Different from the previous research efforts, in-depth data analysis process consists of three levels and is executed in detection and agility components. In particular, level 1 is to generate security objects and related pedigree information from the collected data, level 2 is to determine the Smart Grid system security status with the risk analysis process, and level 3 is to identify the defense strategies related to the current the Smart Grid system status.

Based on the Smart Grid system status, game theoretical based modeling and analysis can be useful to study the interaction between the adversary and the defender. In the game theoretical formalization, the four components are considered: parties, parties' strategies, outcome of the game, and parties' objectives. In our system, we consider two types of parties: attacker and defender. Recall in Section 5.2, we explore the attack space and attacks that can be classified as stealthy attacks and strong attacks. The adversary may have full or partial system knowledge. The adversary's strategy varies based on specific attacks to be launched, such as privilege escalation attack, and malware propagation. Recall in Section 5.4, we investigate effective threat detection schemes. The defenders' strategies rely on these detection

schemes to take corresponding actions, such as intrusion detection system deployment, firewall configuration, and filter configuration. The strategy combination of the defender and attacker determines the outcome of the interaction. Several cases can be considered such as an attack that is launched and not detected; an attack that is launched and detected; the adversary chooses not to launch the attack, but no detection alert is issued; and no attack is launched, but a false alarm is issued. The objectives of adversary are to launch an attack that is hard to be detected and to choose the strategy that maximizes its expected impact. Oppositely, the objectives of the defender are to detect as many attacks as possible and to reduce the number of false positives. Based on the defender's and the adversary's strategies and objectives, the interactions between the adversary and the defender and various game theoretic models can be studied, including static, stochastic, and repeated games.

5.7 Summary

In this chapter, we systematically explored the space of attacks in energy management process and take attacks against distributed energy transmission as an example to investigate the risk of those attacks. Based on the explored attack space, we systematically investigated the defense taxonomy to present how to secure energy management process in the Smart Grid, and proposed defensive strategies. Finally,

we introduced a unified theoretical framework to study the effectiveness of the synergy of risk analysis, threat detection, and defense reactions.

Chapter 6

A Cloud Computing Based Architecture to Improve Efficient and Secured Smart Grid Operations

6.1 Overview

In the Smart Grid, a large amount of data will be collected from physical and cyber components and transmitted to the computing core through communication networks that enables efficient and secured operations of the Smart Grid [147, 148]. For example, in the Smart Grid, renewable energy sources, distributed energy storage, and generation need to be efficiently integrated and managed through complex and computationally intense models, real-time analysis, and visualization. Collected massive streaming data will be generated from power grid to Energy Management System (EMS) to enable efficient system operation [149].

In addition, to provide a highly secured the Smart Grid, a threat monitoring and detection system should be developed to efficiently mitigate cyber-threats against the Smart Grid. Effectively processing of threat monitoring data from both physical and cyber components will facilitate the detection of cyber-threats and help security administrators respond to cyber-threats in a timely manner. Nonetheless, developing a scalable, reliable and robust defense system for the Smart Grid is a challenging task. It is challenging to quantify the impact of threats as they may come from various sources and to detect threats because the detection system has to inspect various

data sources, which are always in large-scale with different formats and semantics. Monitoring the Smart Grid applications (e.g., cyber and physical components) and threat detection are characterized by very high volume data streams and real-time processing requirements. Resources in the Smart Grid (e.g., bandwidth, storage, etc.) are also limited. For example, given a large number of meters in the grid, it becomes a huge computing workload to analyze data and to provide useful information for applications and operators. How to efficiently store and process such big data to assist secured and efficient operation of the Smart Grid has become a challenging and urgent issue.

To address the aforementioned challenges, in this chapter, we study cloud computing based techniques to assist secured and efficient Smart Grid operations (e.g., energy forecast, distributed energy routing, and others). The proposed system consists of data sources cloud infrastructure, and an operation center. Monitoring sensors can be deployed on devices in the Smart Grid to collect the Smart Grid components' information and transmit the information (e.g., raw data or alerts) to the cloud. A cloud infrastructure is a distributed system deployed with a number of servers, providing both storage and computation resources. There are two types of servers in the cloud: storage servers and application servers. The collected streams of data will be pushed and stored in storage servers in real time while application servers will provide data analysis. MapReduce is one type of technique used to

speed up data processing by separating and processing data streams concurrently. Operation center plays the intelligence role that can dynamically update the Smart Grid operation policies and configuration, and monitor the system security.

We leverage the large storage and computing resources in the cloud to conduct big data processing and computation. We investigate and develop a streaming-based storage model to minimize transmission latency and computing delay. To efficiently collect and process large data stream from the Smart Grid components, we investigate stream processing mechanisms (e.g., MapReduce [150]). The MapReduce usually consists of two functions: Map() function and Reduce() function, where Map() function performs filtering and sorting operations on the data and Reduce() function is in charge of summary operation (e.g., finding the number of occurrences of a given pattern) [151].

6.2 MapReduce Framework in Cloud

Generally speaking, the *MapReduce* is a parallel programming model primarily designed for batch processing over big data in a distributed computing environment [74]. Notice the *MapReduce* is designed using the concept of divide-and-conquer and follows the master/slave paradigm. The *MapReduce* can take advantage of the locality of data, processing data on or near the storage assets to reduce the overhead of transmitting data. To address the big data issue, a master node will divide the

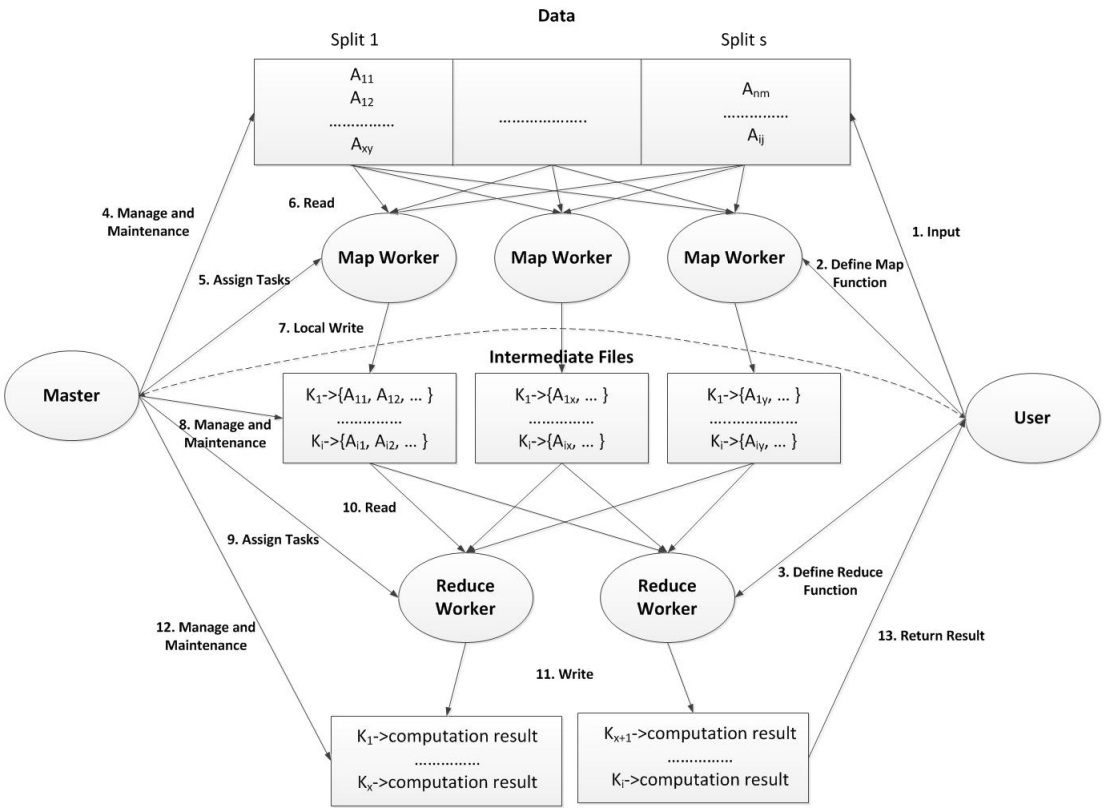


Figure 6.1: A MapReduce Framework

task into a number of small subtasks, which are independently executed in parallel on multiple slave nodes. Slave nodes can be either individual threads, processes, or individual computers. Intermediate results from slave nodes will be further integrated to obtain the final results.

As shown in Figure 6.1, the *MapReduce* framework consists of user, master, map worker, and reduce worker. The user provides a set of data A_{ij} to the *MapReduce* framework for processing. The data set A_{ij} is then split into n chunks and stored in the distributed file system. The *MapReduce* is based on key/value tuples and relies

on two built-in functions: the map function and the reduce function. The map and reduce functions can be defined by the user with a set of key/value pairs $K_i \rightarrow A_{ij}$. The key/vaule pairs can be various data types (e.g. string, integer, etc). For example, the source IP address and the destination IP address pair can be defined as the key with string data type and the information related to the source IP address while the destination IP address pair can be defined as a value. The detail of map and reduce functions will be introduced next.

6.2.1 Map Function

For the map function, after data is input into the *MapReduce* framework, the master will manage and maintain data in the distributed file system. Based on the defined map function with key/value pairs by users, the master will divide the processing task into multiple subtasks and distributes them to map workers. After receiving assigned tasks and data locations, the map worker will read the data from the distributed file system and process data. Then, the map worker will scan the input data and perform key matchings to list associated key/value pairs. The map workers can run subtasks concurrently and the master will keep tracking the progress of individual subtasks. Subtasks in the waiting queue will then be assigned to the map workers when they become available. The output of the map worker will be a set of intermediate key/value pairs $K_i \rightarrow A_{i1} \ldots A_{ij}$, which will be output into intermediate files and stored locally. The intermediate key/value pair is the list of values related

to defined keys. The locations of intermediate key/value pairs will be feedback to the master. In the following, we illustrate an example with the three traffic inputs to the map worker.

```
1. Oct 16 03:07:38 192.168.1.1   192.168.1.2   80
2. Oct 16 03:07:42 192.168.1.23  192.168.1.40  80
3. Oct 16 03:07:42 192.168.1.5   192.168.1.96  23
```

We can see that the traffic input 1 and input 2 are from the same port 80. We define the key as the port number and the value as the recorded information associated with the selected port number. The aggregated results of intermediate key/value pairs are listed as follows:

```
1. 80: ["Oct 16 03:07:38 192.168.1.1   192.168.1.2",
        "Oct 16 03:07:42 192.168.1.23  192.168.1.40" ]
2. 23: ["Oct 16 03:07:42 192.168.1.5   192.168.1.96"]
```

6.2.2 Reduce Function

In the reduce function, prior to the computation, the intermediate results need to be shuffled or sorted to group the identical intermediate key/value pairs located in different intermediate files. Based on the key sorting results $K_1, \ldots, K_i$, the master will assign different tasks to the reduce workers, along with the intermediate key/value pair locations. For example, the intermediate key/value pairs with the key ranging from K_1 to K_x will be assigned to one reduce worker and the intermediate key/value pairs with the key ranging from K_{x+1} to K_i will be assigned to another

reduce worker. The reduce worker will locally or remotely retrieve the intermediate results and perform the key/value computation. The output of the reduce function will be returned to the distributed file system and reported to the master. After all the map and reduce workers complete the assigned subtasks, the master will return final results to the user.

6.2.3 Apache Hadoop

Generally speaking, *Hadoop* [89] is an open-source software framework licensed under the *Apache* v2 license. As a *MapReduce* implementation, *Hadoop* supports data-intensive distributed applications and can work with a number of computation independent computers and deal with petabytes of data. In order to perform the network traffic analysis for cyber security situation awareness, *Hadoop* consists of the following functions:

- *Network Data Capture:* It is responsible for capturing network traffic data. Once network data is captured, this function computes flows and exports data to specified collectors. To capture network traffic data, software tool such as *nProbe* can be used. The traffic capture can be performed by a high-end system with dedicated hardware if needed.

- *Netflow Collection:* It is performed by specific software named collectors such as *nfCapd*, which reads the netflow data collected from the network and stores

it into binary files.

- *Traffic Information Storage:* The software tool such as *nfdump* can be used to read the netflow data from files and dump them to store as plain-text files.

In *Hadoop*, there are several important components such as the distributed file system, the database management system, and the user interface, which all are important to support the above functions. In the following, we will describe these components.

- *HDFS (Hadoop Distributed File System):* It is responsible for data management and manipulation. HDFS provides a reliable storage of both input and output data required by *MapReduce* tasks. In HDFS, data is stored as files that can be split and distributed across multiple nodes. Unlike other distributed file systems, HDFS is explicitly designed for applications with large datasets. It features a high fault-tolerance through data replication, concurrent access to files, cross platform portability, and low cost deployment.

- *Hbase:* It is an open source, non-relational, distributed, and column-oriented database management system that runs on top of HDFS. To enable scalable parallel processing of data, *Hadoop* integrates a tool named *Cloudera Impala*, an open source Massively Parallel Processing (MPP) query engine. *Cloudera Impala* enables the capability for users to issue low-latency SQL queries to

the data stored in HDFS and Apache Hbase without the physically moving or transforming the data.

- *Pig and Hive:* To interact with users, HDFS provides two non-programmatic interfaces *Pig and Hive* to process queries from users and present datasets in a standardized way. *Pig and Hive* receive queries from users, compile queries, and execute them on nodes. *HiveQL* is a query language provided by the hive interface. Similar to the well-known Structured Query Language (SQL), *HiveQL* presents data as tables to perform the basic SQL operations such as select, join, insert, and etc.

6.3 System Architecture

In this section, we first show the system architecture, which consists of several key components: data sources, cloud infrastructure, and an operation center.

Data Sources: The components in the Smart Grid consists of physical devices (e.g., smart meter, storage devices, and others), end hosts (e.g. computers, mobile devices, and others) and network devices (e.g., routers, firewalls, and others). As the resources of components are limited, they lack the computation ability and storage capacity in comparison with high performance computers. The cloud infrastructure will provide a huge storage capacity and computation power to perform efficient and secured energy management. Hence, the data from distributed

the Smart Grid devices are continuously streamed up to the cloud infrastructure for monitoring and analysis purposes. For example, for the Smart Grid security management, various logs (e.g., system logs, security logs, application logs, and etc.) on physical devices and computers can be used to perform both static and dynamic behavior analysis of malware on the Smart Grid components [152]. With the help of network devices (e.g., routers, firewalls, sniffers, etc.), the Smart Grid network traffic data (e.g. the number of scans from designated sources and destinations, etc.) can be collected and used to detect attacks. To provide an efficient and fast data retrieval with processing for threat detection, the collected data shall be normalized in a specific format.

Cloud Infrastructure: The cloud computing infrastructure is composed of multiple distributed servers, which are responsible for provisioning storage and computing resources to cyber security applications. Pushing storage, computation and analysis to the cloud will not only resolve the issue of the limited resources on an Smart Grid device, but also significantly improve the efficient and secured energy management through the fast data retrieval and processing. More importantly, the implementation of a large data processing technique such as the *MapReduce* will make the system more efficient through eliminating operation delays and enabling real-time processing of data streams. In addition, the reliability of the system can be improved as servers in the cloud infrastructure are immune to the single node of

failure.

Operation Center: In the Smart Grid, the operation center can hold the intelligent role and be responsible for analyzing the stream data stored in the cloud to perform efficient and secured energy management. The operation center interacts with distributed grid devices and cloud servers by pushing cyber operation policies and configuration to the monitored end-user devices and cloud servers. To compound this system, the data visualization will provide the operation center to real time power consumption situations and conduct the cyber security situation awareness to deal with emergent and dangerous cyber threats.

6.4 A Cloud Computing Based Architecture to Improve Efficient Smart Grid Operations

To make the system efficient, we then introduce the two key function modules in our system: data storage module and task scheduling module.

6.4.1 Data Storage Module

In the Smart Grid, the data streams can be collected from the Smart Grid components for the efficient and secured energy management. However, to meet real-time requirements posed by the efficient and secured energy management, there are several challenges. First, a large number of monitoring data streams, including both

host-based and network-based data, is collected from the Smart Grid devices distributed over the network and streamed to the central database for detection and analysis purposes. The mounting volume of data stored in the central database and the continuously increasing storage capacity incur a high cost and can significantly slow data extraction and the overall performance of the system. Second, the real-time data processing to generate useful information is time critical and the latency should be kept at a minimum in order to assure the effectiveness of energy management.

To address these challenges, we can leverage the large storage and computational resources in the cloud to conduct the efficient and secured energy management. To do so, a streaming-based storage model is developed in order to reduce the storage processing time. The cloud infrastructure consists of a number of cloud servers distributed across multiple locations to accommodate threat monitoring and detection on dispersed end user devices in the enterprise network. The cloud storage server provides bi-directional data synchronization capability to distributed devices, reducing the time needed for storage, and avoiding a potential bottleneck of a centralized system. Additionally, the streaming-based storage model should consider the tradeoff between transmission delay incurred by the available bandwidth, and the propagation delay incurred by the distance between user devices and the storage servers.

In the cloud infrastructure, a number of cloud storage servers are distributed across multiple locations. The data stream can be transmitted to the storage server based on an optimal route, reducing the time needed for storage. Distributed cloud storage servers can store data based on the locations of end-user devices over time. With the contiguous data stream stored in the cloud, the status of storage server can be formulated as

$$\int_{t_n}^{t_{n+1}} B_j(x)dx + Q_j(t_n) = Q_j(t_{n+1}), \tag{6.4.1}$$

where $0 \leqq Q_j(t_{n+1}) \leqq \delta \cdot Q_j^{max}$, δ is alert threshold and Q_j^{max} is the maximum capacity of storage server j. When the $Q_j(t_{n+1}) > \delta \cdot Q_j^{max}$, the alert will be issued and broadcasted to the application servers if there is a little storage space on server j.

To rapidly and efficiently store a large amount of data and minimize the delay of storage process, we need to consider the following constraints: (i) The delay for the big data storage mainly consists of a data propagation delay and a transmission delay; (ii) The data propagation delay is affected by the distance $L_{ij}(t)$) between the end-user device i ($i \in [1,m]$) and the cloud storage server j ($j \in [1,n]$; (iii) The data transmission delay relies on the available link bandwidth $B_{ij}(t)$; (iv) The sum of supported bandwidth to each device cannot exceed the total available link bandwidth B_j on the storage server; (v) The size of data in the storage server j cannot exceed the maximum storage capacity; (vi) The dataset from a user device i shall

be split at least S_i chunks and stored in different storage servers; (vii) All the distributed stored data chunks associated with the user device i can restore the whole dataset associated with the user device i.

Based on the above constraints, we formalize the optimal data storage process as an optimization problem, which is listed as follows:

$$\textbf{\textit{Objective.}} \quad Min \left\{ \sum_{i\in[1,m]} (T_i^1 + T_i^2) \right\} \tag{6.4.2}$$

$$\textbf{\textit{S.t.}}$$

$$\begin{cases} \forall i \in [1,m], \forall j \in [1,n], \\ \sum\limits_{j\in[1,n]} (\alpha \cdot L_{ij} \cdot x_{ij}) = T_i^1, \\ \sum\limits_{j\in[1,n]} (\frac{d_{ij}}{B_{ij}} \cdot x_{ij}) = T_i^2, \\ \sum\limits_{i\in[1,m]} (x_{ij} \cdot B_{ij}) \leq B_j, \\ \sum\limits_{i\in[1,m]} (x_{ij} \cdot d_{ij}) + Q_j(t) \leq Q_j^{max}, \\ \sum\limits_{i\in[1,m]} x_{ij} \geq S_i, x_{ij} \in \{0,1\}, \\ \sum\limits_{j\in[1,n]} (x_{ij} \cdot d_{ij}) = d_i, \end{cases}$$

where α is constants, d_i is the size of data collected from user device i, and d_{ij} is the data chunk size of user device i. When the network traffic is low, few data streams share the bandwidth concurrently, so that the propagation delay T_i^1 plays a key role in the total delay of the storage process. Hence, using the shortest path to route the traffic can incur the smallest propagation delay because L_{ij} is the main factor. When the network traffic is higher, the transmission delay incurs a larger impact

on the total delay of the storage process in comparison with the prorogation delay. The idle or not fully loaded storage server will be selected to receive data in order to reduce the delay in the data transmission process. Hence, the storage server selection should be determined by the available bandwidth B_{ij} to the user device.

Nonetheless, the computation overhead for the optimization increases rapidly as the increase in scale of cloud. To overcome this, we consider dividing the large network into multiple regions based on the locations of the Smart Grid components. The optimal data storage process can be performed in each region and then the optimal data storage process can be performed across multiple regions.

6.4.2 Task Scheduling Module

When there is a large amount of data stored in the cloud, how to effectively schedule the data processing is a challenging issue. To this end, an optimal task scheduling model should be developed to assign tasks optimally to slave nodes with an objective of achieving the shortest data processing time. To minimize the data processing time, all the tasks shall be balanced to different servers that perform the computation. In an ideal condition, when all the servers concurrently complete the computations of tasks, the time taken to process all tasks shall be the shortest. Nonetheless, in reality, all the servers cannot complete all tasks computations simultaneously.

As such, the problem of the optimal task scheduling can be formalized with

considering the following constraints: (i) The objective is to minimize the variance of data processing time on all servers; (ii) The data processing time of each server is affected by the server computation speed, data set size, and available bandwidth of the server; (iv) The server computation speed is related to the status of CPU and memory; (v) Each data shall be processed in one server. With the aforementioned constraints, the optimal task scheduling model can be formalized as follows:

$$\textbf{\textit{Objective.}} \quad Min \left\{ Max\{ \sum_{s\in[1,z]} (T_s - \overline{T})^2 \} \right\} \tag{6.4.3}$$

$$\textbf{\textit{S.t.}}$$

$$\begin{cases} \forall k \in [1,w], \forall s \in [1,z], \\ \sum\limits_{s\in[1,z]} T_s \cdot \frac{1}{z} = \overline{T}, \\ \sum\limits_{k\in[1,w]} [x_{ks} \cdot (\frac{d_k}{v_s} + \frac{d_k}{B_s})] = T_s, \\ \omega_1 M_s + \omega_2 C_s = v_s, \\ \sum\limits_{s\in[1,z]} x_{ks} = 1, x_{ks} \in \{0,1\}, \end{cases}$$

where w is the number of tasks, z is the number of servers, T_s is the data processing time on the server s, $\overline{T}$ is the average data processing time on all servers, d_k is the data size of the task k, B_s is the available bandwidth of the server s, v_s is the computation speed of the server s, M_s is the memory status of the server s, C_s is the CPU status of the server s, and ω_1 and ω_2 are the weight coefficients. Based on the modeling, we can see that when $\overline{T} = T_1 = T_2 \ldots = T_z$, the objective function can be minimized such that all the servers and all the tasks can be computed in the

shortest time.

6.5 A Cloud Computing Based Architecture to Improve Smart Grid Security

In this section, we will introduce the cloud based threat detection and then discuss the scene investigation based on the system architecture we proposed.

6.5.1 Threat Detection

To improve the operational efficiency and reliability of power grids through the use of cyber computing and communication technologies, the risks of cyberspace breach on power grid systems need to be seriously studied before massively deploying the Smart Grid technologies. There have been growing concerns in the Smart Grid on protection against malicious cyber threats. The operation and control of the Smart Grid depend on a complex cyberspace of computers, software, and communication technologies. Because the measurement component supported by smart equipment (e.g., smart meters and sensors) plays an important role, it can be a target for attacks. Because those measuring devices may be connected through open network interfaces and lacking tamper-resistance hardware increases the possibility to be compromised by the adversary. The adversary may modify data and compromise measuring components by injecting malicious codes into the memory of measuring components

To address the challenging issue of detecting cyber threats through a large number of data streams in the Smart Grid, we investigate the cloud based threat detection using parallel computing techniques. As an example, signature-based threat detection relies on a set of rules that are used to match packets. Rules can be used to represent characteristics of known attacks, such as the protocol type, port number, packet size, packet content (both strings and regular expressions), and the position of the suspicious content. Both signature-based detection and anomaly-based detection will be integrated to systematically examine system logs, configurations, and traffic logs stored in the cloud.

The cloud based threat detection consists of two sub-components: a parallel signature-based detection module and an anomaly-based detection module. First, the parallel signature-based detection will be used at the front-end of the cloud to detect known external threats through the signatures string matching on captured threat information. A traffic splitter balances and distributes workload to MapReduce slave servers. Each MapReduce slave server will conduct the string matching on threat signatures and report the result. Second, the anomaly-based detection module conducts anomalies analysis on logs stored in the cloud to uncover potential intrusion, misuse, and abnormal behaviors.

To improve detection efficiency, we will develop a MapReduce based machine learning scheme to deal with big threat monitoring data efficiently. The main idea

is to speed up the machine learning process using cloud computing and MapReduce. The first step is to collect the characteristics of threat monitoring data from threat monitors in the Smart Grid. To accurately and rapidly detect anomalies, techniques (e.g., MapRduce based machine learning (MML) schemes, etc.) can be used to profile the dynamic characteristics of collected threat monitoring data and then used to detect anomalies. In this way, the computational burden of the learning process is spread across multiple machines and the learned computational results from multiple machines are then integrated into one single learned classifier. Lastly, the learned classifier will be used to recognize whether a new observed data is normal or abnormal.

6.5.2 Attack Scene Analysis

To improve the Smart Grid security, it is critical to determine the temporal and spatial characteristics associated with attacks. To this end, we can conduct attack scene analysis in both temporal and spatial domains to understand the attack intent and behaviors. The identified attack information will be visualized to provide useful information to the human analyst.

The spatiotemporal correlation technique can be adopted to correlate events across both spatial and temporal domains, and identify aggregated abnormal event patterns and security status of monitored end user devices. Generally speaking, correlation is defined as establishing or finding a relationship between entities. It is

a recognized technique in cyber security to improve the effectiveness of threat detection and analysis process by combining information from multiple sources. The spatiotemporal correlation technique can enable the capture of abnormal patterns and malicious behavior by analyzing a sequence of events that occur in different time sequences and locations across multiple segments in the network. An event is a sequence of actions that, directly or indirectly, cause the security state of end user devices to transit from one to another. A number of salient features should be considered in the system such as identifying the origins of attacks, circumscribing the location and the region of occurrence, identifying potential compromised devices based on their proximity to the regions under attack, predicting the magnitude of the attack, and predicting future attacks.

In addition, to visualize the threat data and provide more meaningful information, we can develop visualization features, including 1D, 2D and 3D (geolocation-based and Google map-based) displays. To effectively detect attacks, both the discrete wavelet transform (DWT) based approach and the traffic volume based approach can be used to transform the real-world information in IP addresses and other detection features into images and investigate the patterns of threats. The large number of IP addresses to retrieve the results for display will incur a high computation overhead. Hence, the MapReduce paradigm can help map the data into multiple tasks based on the time sequence and/or network regions and assign

each task to an individual slave server. Slave servers will then compute the visualization of a partial system and pass the results back to the master server, where the information will be further aggregated and stacked to produce a global visualization.

6.6 Implementation

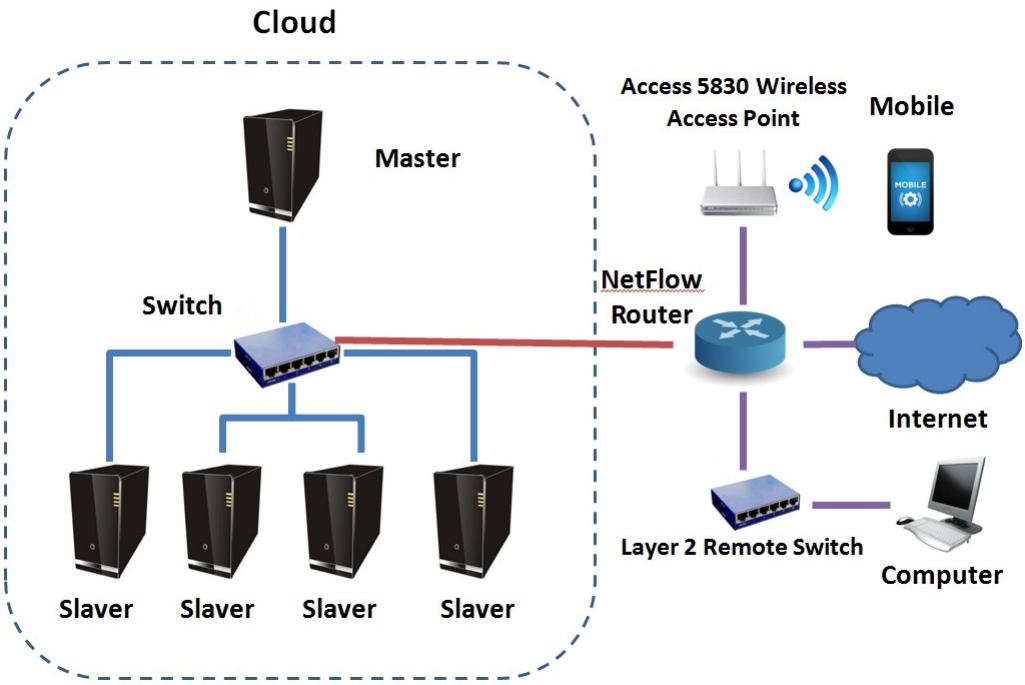


Figure 6.2: System Testbed

We implemented the cloud computing based system for efficient and secured energy management, which consists of one master node and four slave nodes, as shown in Figure 6.2. The slave node can support both the map worker and reduce workers. In each node, DELL Optiplex 9010 computer with Intel Core i7 3.40GHZ

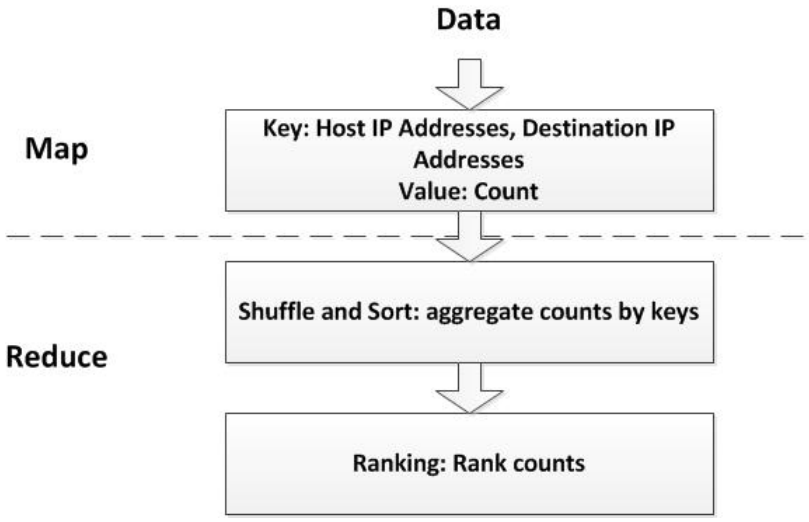


Figure 6.3: Ranking Workflow

8 processors and 16GB RAM and 2TB hard drive is used. We use the *Cloudare Manager* as a single and central interface to carry out the configuration, management, and monitoring of the designed system.

We downloaded the *cloudera-manager-installer.bin* from the *Cloudera* website. We configure it with the executable permission by the command *"chmod u+x cloudera-manager-installer.bin*" and the installer can be executed with the command *"sudo ./cloudera-manager-installer.bin*" to install the *Cloudera Manager*. The server is set on port 7180. To install *Cloudera Manager* on hosts in the cloud, the *Cloudera Manager Admin Console* is used to install and configure CDH (Cloudera Distribution including Apache Hadoop). It is worth noting that CDH is an open source, powerful management, and automation tool to deploy Apache Hadoop widely.

Our developed system workflow consists of four steps: data collection, data

normalization, data computation, and data visualization. We use the scenario of analyzing network traffic data to find anomalous behavior as an example to show how our developed prototypical system performs cyber security management in the Smart Grid and use the scenarios of analyzing energy consumption to demonstrate the efficiency of energy management improvement.

- *Step 1: Data Collection:* The agent is distributed in the network to collect the useful data and monitor the activities from distributed devices. It is worth noting that the collected data is unstructured and unreadable, which needs to be normalized. For example, the collected traffic data is stored in the network traffic trace files with the *PCAP* format, which is binary and not readable.

- *Step 2: Data Normalization:* The collected data is normalized and stored in HDFS. To filter out the useless information and normalize the collected data, we use a tool called *tshark*, which is a wireshark network analyzer, which integrates the packet capture function and the *PCAP* format data reading function. *tshark* normalizes the unstructured data into the structured data with a specific format. For example, the network traffic data can be normalized based on the characteristics of frame number, source IP address, destination IP address, and port number.

- *Step 3: Data Computation:* In this step, the normalized data is computed using the *MapReduce* framework.

- *Step 4: Data Visualization:* To help the Smart Grid administrator to detect anomalies in the network and present useful information to defend against cyber attacks, the data visualization is implemented to transform the data such as IP addresses and other features (e.g. time, ports, and others) into images, which can better present the attack consequence and scenes.

6.7 Performance Evaluation

In this section, we will present the effectiveness of cloud computing architecture to assist efficient energy management and security in the Smart Grid.

6.7.1 Improving Efficiency of Energy Management

Table 6.1: Data Range and Time Scale

Data Type	**Range**
ID of Houses	1-283
Time Interval	Hourly
Time Span	Approximately 200 days
Number of Data Points	Approximately 4800 (one per hour)

Table 6.2: Data Fields

Max_Temp	Mean_Temp	Min_Temp
Max_WindSpeed	Mean_WindSpeed	ID
Day-of-Year	Hour	Electricity Consumption

To improve efficient energy management, we take a scenario of analyzing energy consumption of each house in a time interval. We use the real-world data set

Table 6.3: Sample of House Information

ID	Building	Rent	Year Const.	Size
1001	Townhouse, duplex or row house	Rent	2004	92.90-139.35 sq. meters
1002	Single Family Detached House	Own	1992	185.81-232.37 sq. meters

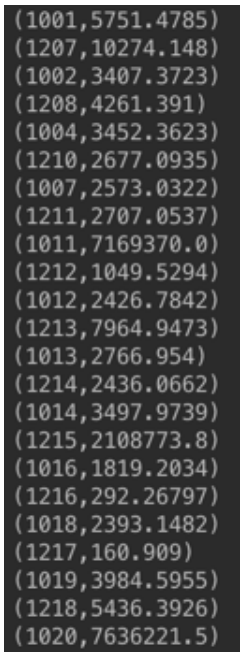


Figure 6.4: Energy Consumption Result (House ID, Energy Usage)

from Stanford University, which consists of meter readings from 283 houses over 200 days (between February 2010 and October 2010) [153]. The size of this data set is 46.3MB. To demonstrate the effectiveness of cloud computing in the Smart Grid, we duplicate the data set to create multiple data sets with different data sizes (e.g., 30GB, 60GB, 90GB, and 120GB). An example of meter reading is shown in Table 6.1. From the table, each house is assigned an ID. The meter reading data for energy usage is measured hourly. The fields contained in the data set are shown

in Table 6.2, which consists of the house ID , time, energy usage, the maximum, mean, and minimum value of temperature, and maximum and mean value of wind speeds. The house size (i.e., the area) is included as well. As an example, the information shown in Table 6.3 is the data associated with house 1001 that is for a rented townhouse, built in 2004, with $92.90 - 139.35$ sq. meters.

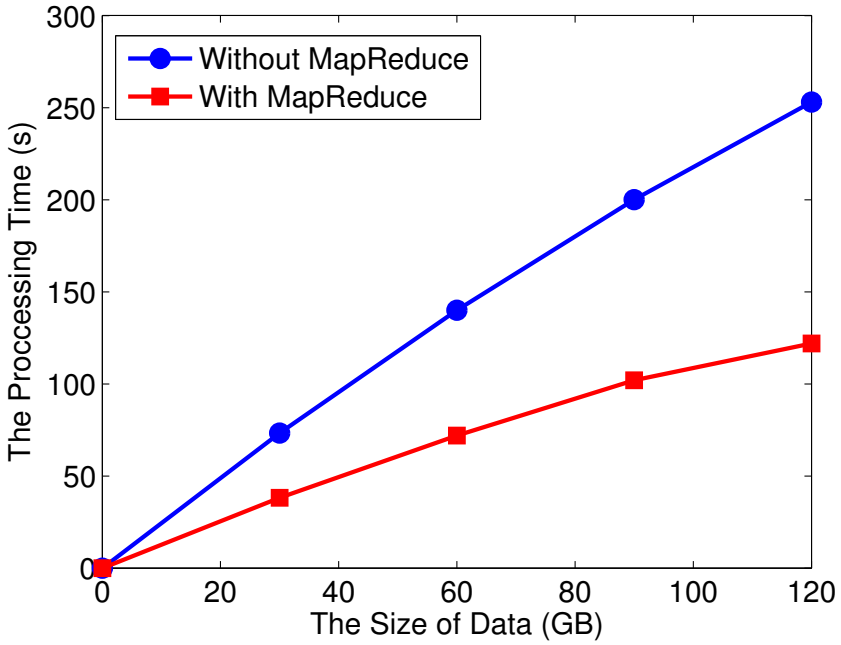


Figure 6.5: The Processing Time versus The Size of Data

Figure 6.4 shows the result of energy consumption for each house in 200 days. For example, the house ID 1001 totally consumes 5751.4785KWh in 200 days. As shown in Figure 6.5, comparing the computation of energy consumption with MapReduce and without MapReduce, we can see that MapReduce can achieve fast and efficient data process. The processing time increases with the increase of data

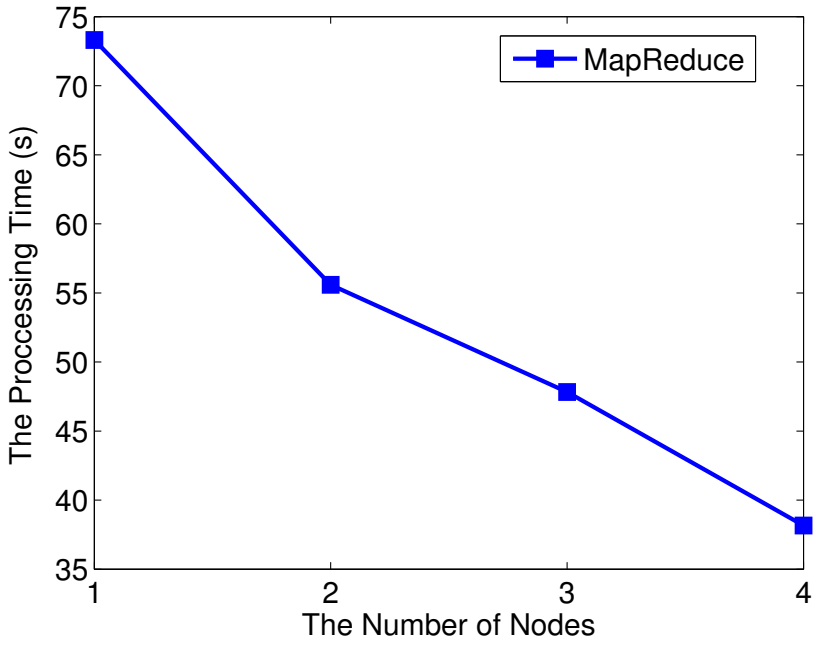


Figure 6.6: The Processing Time versus The Number of Nodes

size. When more slave nodes are applied in the computation process, as shown in Figure 6.6, the processing time will significantly decrease.

6.7.2 Improving Security in the Smart Grid

To evaluate the effectiveness of cloud computing to improve cyber security in the Smart Grid, we use the network traffic data from *http://www.caida.org/home/* to perform experiments. The data consists of a 160.8 GB data file and split into 3 GB per chunk. In our evaluation, we consider two representative scenarios: ranking and aggregation, as examples to demonstrate the effectiveness of our developed system.

To identify whether or not distributed devices have high scan traffic rates, we implemented the ranking primitive, which has been widely used for security management. Figure 6.3 illustrates the workflow of the ranking primitive implementation in the *MapReduce* framework, where the host IP address and destination IP address pair is defined as the key and the count as value. Users define the map function to list the count for the key and the reduce function to rank the count. Based on the map function and the reduce function defined by users, the master will schedule tasks to the map worker and track the status of tasks. With the assigned tasks, the map worker will read the data from the distributed file system and match the data to the defined key when the count is recorded. The shuffling and sorting operation will group all counts with the same key in intermediate results and the sort results are based on the key. Then, the reduce worker will count the ranking results to identify the host IP addresses, which have the high scan traffic rate. Figure 6.7 illustrates the result for conducting port ranking. As we can see, the port 80 is most scanned, which is about 138 millions.

In cyber security management in the Smart Grid, the aggregation is another important primitive, which can be used to consolidate the analyze results and conduct statistical analysis, as a critical part of intrusion detection. The traffic data features (e.g., time, source address, destination address, port number, etc.) can be used as the key to perform the aggregation. Using the port number as the aggregation key,

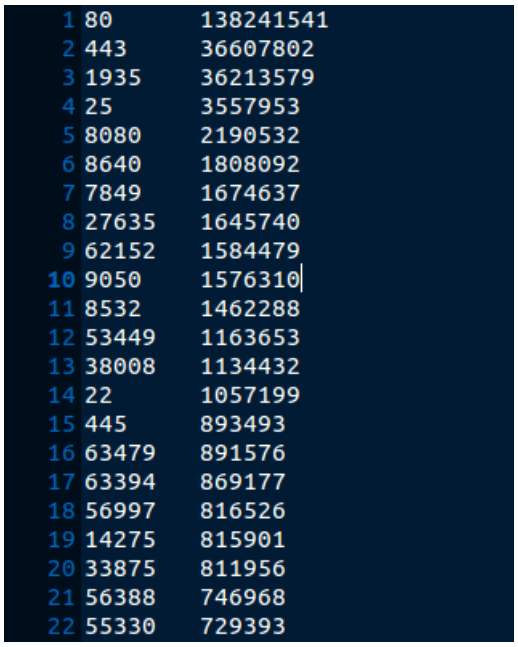


Figure 6.7: Ports Ranking Result

the information related to the key will be defined as the value, which presents the volume of data associated with the designated port. The workflow of traffic aggregation is similar to the ranking in the *MapReduce* framework. The large amount of network traffic data will be split and assigned to the map worker. The map worker finds the information related to the port number and lists them. The reduce worker will then aggregate all the related information with the same key and then generate the final aggregation results.

Figure 6.8 shows an screenshot of the ranking in the *MapReduce* framework evaluation result. We summarize all the evaluation results in Table 6.4. We can see that in order to process 160.8 GB of data, the processing speed of the ranking and

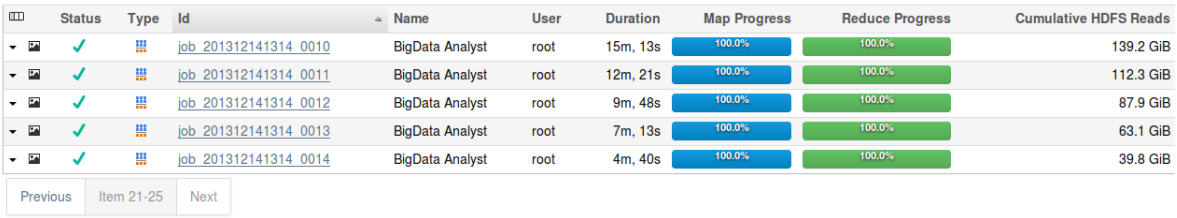
Table 6.4: Time versus The Size of Data

		Processing Time (s)			
		Ranking with MapReduce	Ranking without MapReduce	Aggregation with MapReduce	Aggregation without MapReduce
Data Size (GB)	39.8	280	1401.76	205.44	1047.53
	63.1	433	2222.382	333.46	1660.79
	87.9	598	3095.8	460.35	2313.5
	112.3	741	3930.5	588.14	2955.7
	139.2	913	4902.6	728.02	3663.7
	160.8	1043	5663.4	842.15	4232.2

Table 6.5: Time versus The Number of Slave Nodes

		Processing Time (s)	
		Ranking with MapReduce	Aggregation with MapReduce
The Number of Slave Nodes	1	4200	3346.7
	2	2061	1644.69
	3	1364	1094.21
	4	1043	842.15

the aggregation can be improved after the *MapReduce* framework is introduced. With the increase of data size, the time for processing data will increase as well. Table 6.5 presents the relationship between the processing time and the number of slave nodes. As we can see, with more slave nodes used in the system, more computation resources can be applied to process data, leading to a declined trend of data processing time. For example, to process 160.8 GB of data for the ranking and the aggregation using four slave nodes, the processing time are 1043 seconds and 842.15 seconds, respectively.



Status	Type	Id	Name	User	Duration	Map Progress	Reduce Progress	Cumulative HDFS Reads
✓		job_201312141314_0010	BigData Analyst	root	15m, 13s	100.0%	100.0%	139.2 GiB
✓		job_201312141314_0011	BigData Analyst	root	12m, 21s	100.0%	100.0%	112.3 GiB
✓		job_201312141314_0012	BigData Analyst	root	9m, 48s	100.0%	100.0%	87.9 GiB
✓		job_201312141314_0013	BigData Analyst	root	7m, 13s	100.0%	100.0%	63.1 GiB
✓		job_201312141314_0014	BigData Analyst	root	4m, 40s	100.0%	100.0%	39.8 GiB

Previous Item 21-25 Next

Figure 6.8: Result Screenshot

To present useful information to security administrators, we also implemented the network traffic visualization using the *MapReduce* framework as shown in Figure 6.9. Here, the horizontal axis represents the source IP address and the vertical axis represents the destination IP address. The data point in this figure represents the communication density of a pair of nodes. As we can see, the address 0.3.143.223 has a high scan traffic and is highly likely to be a worm propagation host.

6.8 Summary

In this chapter, we investigated a cloud computing based system for CPS operations to increase the capacity for large data. Using the *MapReduce* framework, we designed and implemented a system consisting of data sources, cloud infrastructure, and an operation center. To make our proposed system efficient, we introduced two key function modules of our system: data storage module and task scheduling module. We performed the *MapReduce* implementation using the *Apache Hadoop* and the implemented system consists of data collection, data normalization, data

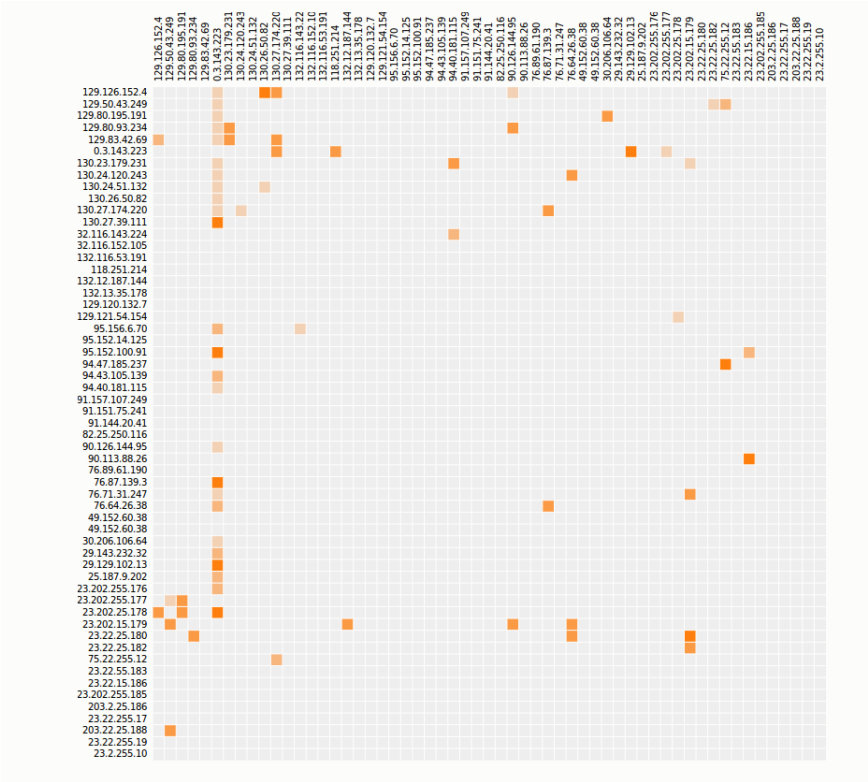


Figure 6.9: Visualization

computation, and data visualization. Using ranking and aggregation as examples of primitive functions for security management and energy consumption computation for energy management, we carried out experiments and our data shows that our system can efficiently analyze a large amount of data.

Chapter 7

Concluding Remarks

In this dissertation, we developed a framework to enable efficient and secured energy based CPS. To be specific, we first developed schemes to distribute energy resources as well as storage devices and developed schemes to address the uncertainty from both cyber and physical components. We established a uniformed framework to model and analyze the effectiveness of those schemes and consider the interactions between cyber and physical components. We then systematically explored the space of attacks in the energy management and investigated the risk of those attacks. We also discussed countermeasures (e.g., prevention, detection and response) against those attacks. Finally, we proposed a uniformed cloud computing based architecture to assist assist efficient and secured the Smart Grid operations.

In the near future and beyond, I plan to research other CPSs and big data problems that I believe are novel, challenging, important, and impactive to the community, such as efficiency and security in intelligent transportation systems, and security in big data. My current research provides me with a solid and unique foundation to carry out my proposed future research.

References

[1] Cyber-physical system. [Online]. Available: http://en.wikipedia.org/wiki/Cyber-physical_system

[2] Smart grid. [Online]. Available: http://energy.gov/oe/services/technology-development/smart-grid

[3] Nist and the smart grid. [Online]. Available: http://www.nist.gov/smartgrid/nistandsmartgrid.cfm

[4] T. Basso, J. Hambrick, and D. DeBlasio, "Update and review of ieee p2030 smart grid interoperability and ieee 1547 interconnection standards," in *Proccedings of IEEE International Conference on Innovative Smart Grid Technologies (ISGT)*, 2012, pp. 1–7.

[5] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[6] Hurricane sandy. [Online]. Available: https://en.wikipedia.org/wiki/Hurricane_Sandy

[7] J. G. Kassakian, R. Schmalensee, G. Desgroseilliers, T. D. Heidel, K. Afridi, A. Farid, J. Grochow, W. Hogan, H. Jacoby, J. Kirtley *et al.*, "The future of the electric grid," *Technical Report in Massachusetts Institute of Technology*, 2011.

[8] D. Zhang, L. Ge, W. Yu, H. Zhang, R. L. Hardy, and R. J. Reschly, "On effective data aggregation techniques in host–based intrusion detection in manet," *International Journal of Security and Networks*, vol. 8, no. 4, pp. 179–193, 2013.

[9] A. Molderink, V. Bakker, M. G. Bosman, J. L. Hurink, and G. J. Smit, "Management and control of domestic smart grid technology," *IEEE transactions on Smart grid*, vol. 1, no. 2, pp. 109–119, 2010.

[10] T. Znati, "Security for emerging cyber-physical systems research challenges and directions," in *Proccedings of First International Workshop on Data Security and PrivAcyin wireless Networks Panel*, 2010.

[11] I. Lee, "Assuring the safety, security and reliability of medical device cyber physical systems," in *Proccedings of NSF CPS*, 2012.

[12] Cyber-physical systems week. [Online]. Available: http://www.cpsweek.org/

[13] M. D. Ilic, L. Xie, U. Khan, and J. Moura, "Modeling future cyber-physical energy eystems."

[14] L. Sha, S. Gopalakrishnan, X. Liu, and Q. Wang, "Cyber-physical systems: A new frontier," *Journal of Machine Learning in Cyber Trust*, pp. 3–13, 2009.

[15] L. Parolini, N. Tolia, B. Sinopoli, and B. H. Krogh, "A cyber-physical systems approach to energy management in data centers," in *Proceedings of the 1st ACM/IEEE International Conference on Cyber-Physical Systems*, 2010, pp. 168–177.

[16] F. Zhang and Z. Shi, "Optimal and adaptive battery discharge strategies for cyber-physical systems," in *Proceedings of the 48th IEEE Conference on Decision and Control*, 2009, pp. 6232–6237.

[17] E. A. Lee, "Cyber-physical systems are computing foundations adequate," in *Position Paper for NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap*, vol. 2, 2006.

[18] Roadmap for smart grid interoperability standards. [Online]. Available: http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pdf

[19] K. Wan, K. Man, and D. Hughes, "Specification, analyzing challenges and approaches for cyber-physical systems (cps)," *Journal of Engineering Letters*, vol. 18, no. 3, p. 308, 2010.

[20] Cps submit report. [Online]. Available: http://varma.ece.cmu.edu/Summit/

[21] M. Xue, S. Roy, Y. Wan, and S. K. Das, "Security and vulnerability of cyber-physical," *Handbook on Securing Cyber-Physical Critical Infrastructure*, p. 5, 2012.

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

[23] E. K. Wang, Y. Ye, X. Xu, S. Yiu, L. Hui, and K. Chow, "Security issues and challenges for cyber physical system," in *Proceedings of the 2010 IEEE Internatinal Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*, 2010, pp. 733–738.

[24] Q. Shafi, "Cyber physical systems security: A brief survey," in *Proccedings of 12th IEEE International Conference on Computational Science and Its Applications*, 2012, pp. 146–150.

[25] Q. Zhu, C. Rieger, and T. Bacsar, "A hierarchical security architecture for cyber-physical systems," in *Proccedings of 4th IEEE International Symposium on Resilient Control Systems (ISRCS)*, 2011, pp. 15–20.

[26] X. Jin, A. Ray, and R. M. Edwards, "Integrated robust and resilient control of nuclear power plants for operational safety and high performance," *IEEE Transactions on Nuclear Science*, vol. 57, no. 2, pp. 807–817, 2010.

[27] E. M. Atkins, "Cyber-physical aerospace: Challenges and future directions in transportation and exploration systems," in *National Science Foundation Workshop on Smart Transportation and Aviation*, 2006.

[28] P. Antsaklis, "From hybrid to networked cyber-physical systems," in *Proccedings of IEEE International Conference on American Control Conference*, 2009, pp. 5804–5805.

[29] D. Shen, G. Chen, K. Pham, and E. Blasch, "A trust-based sensor allocation algorithm in cooperative space search problems," in *Proccedings of International Conference on SPIE Defense, Security, and Sensing*, 2011, pp. 80 440C–80 440C.

[30] A. Banerjee, K. K. Venkatasubramanian, T. Mukherjee, and S. K. S. Gupta, "Ensuring safety, security, and sustainability of mission-critical cyber–physical systems," *Journal of Proceedings of the IEEE*, vol. 100, no. 1, pp. 283–299, 2012.

[31] Modeling, simulation, information technology & processing roadmap. [Online]. Available: http://www.nasa.gov/pdf/501321main_TA11-MSITP-DRAFT-Nov2010-A1.pdf

[32] Control systems cyber security: Defense in depth strategies. [Online]. Available: http://www5vip.inl.gov/technicalpublications/Documents/3375141.pdf

[33] D. Li, G. Chen, E. Blasch, and K. Pham, "Sensor attack avoidance: linear quadratic game approach," in *Proccedings of IEEE International Conference on Information Fusion*, 2009, pp. 1131–1138.

[34] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[35] A. Cardenas, S. Amin, G. Schwartz, R. Dong, S. Sastry *et al.*, "A game theory model for electricity theft detection and privacy-aware control in ami systems," in *Proccedings of 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 1830–1837.

[36] C. Y. Ma, N. S. Rao, and D. K. Yau, "A game theoretic study of attack and defense in cyber-physical systems," in *Proccedings of IEEE International Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2011, pp. 708–713.

[37] Q. Zhu and T. Başar, "Robust and resilient control design for cyber-physical systems with an application to power systems," in *Proccedings of IEEE International Conference on Decision and Control and European Control Conference (CDC-ECC)*, 2011, pp. 4066–4071.

[38] R. Mitchell and I.-R. Chen, "A hierarchical performance model for intrusion detection in cyber-physical systems," in *Proccedings of IEEE International Conference on Wireless Communications and Networking Conference (WCNC)*, 2011, pp. 2095–2100.

[39] W. Yu, X. Fu, E. Blasch, K. Pham, D. Shen, G. Chen, and C. Lu, "On effectiveness of hopping-based spread spectrum techniques for network forensic traceback," in *Proccedings of IEEE14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2013, pp. 101–106.

[40] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Proccedings of ACM International Conference on the 47th Design Automation Conference*, 2010, pp. 731–736.

[41] B. McMillin, C. Gill, M. Crow, F. Liu, D. Niehaus, A. Potthast, and D. Tauritz, "Cyber-physical systems distributed control: the advanced electric power

grid," in *Proccedings of IEEE International Conference on Electrical Energy Storage Applications And Technologies*, 2007.

[42] A. J. Conejo, J. M. Morales, and L. Baringo, "Real-time demand response model," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 236–242, 2010.

[43] X. Guan, Z. Xu, and Q.-S. Jia, "Energy-efficient buildings facilitated by microgrid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 243–252, 2010.

[44] Career: Systematic multi-scale integration of physics-based and data-driven models of distributed resources for enabling ubiquitous energy storage services in power systems. [Online]. Available: http://www.nsf.gov/awards/award_visualization_noscript.jsp?org=ECCS &region=US-TX&instId=0036327060

[45] Research evaluation of wind generation, solar generation,and storage impact on the california grid. [Online]. Available: http://www.energy.ca.gov/2010publications/CEC-500-2010-010/CEC-500-2010-010.PDF

[46] Variability of wind power and other renewables - management options and strategies. [Online]. Available: http://www.uwig.org/iea_report_on_variability.pdf

[47] Matching hourly and peak demand by combining different renewable energy sources: A case study for california in 2020. [Online]. Available: http://web.stanford.edu/group/efmh/jacobson/Articles/I/CombiningRenew /HosteFinalDraft

[48] J. Taneja, R. Katz, , and D. Culler, "Defining cps challenges in a sustainable electricity grid," in *Proccedings of IEEE International Conference on Cyber-Physical Systems*, 2012.

[49] M. D. Ilic and L. Xie and J. Y. Joo, "Efficient coordination of wind power and price-responsive demand-part ii: Theoretical foundations," *IEEE Transactions on POWER SYSTEMS*, vol. 26, no. 4, pp. 1885–1893, 2011.

[50] A. Evans, V. Strezov, and T. Evans, "Assessment of utility energy storage options for increased renewable energy penetration," *Journal of Renewable & Sustainable Energy Reviews*, vol. 16, no. 6, pp. 4141–4147, 2012.

[51] The role of energy storage with renewable electricity generation: Technical report. [Online]. Available: http://www.nrel.gov/docs/fy10osti/47187.pdf

[52] D. Connolly, H. Lund, B. Mathiesen, E. Pican, and M. Leahy, "The technical and economic implications of integrating fluctuating renewable energy using energy storage," *Journal of Renewable Energy*, vol. 43, pp. 47–60, 2012.

[53] Y. Levron and D. Shmilovitz, "Power systems optimal peak-shaving applying secondary storage," *Journal of Electric Power Systems Research*, vol. 89, pp. 80–84, 2012.

[54] A. A. Cardenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and communications security*, 2011, pp. 355–366.

[55] A. Teixeira, G. Dán, H. Sandberg, and K. H. Johansson, "A cyber security study of a scada energy management system: Stealthy deception attacks on the state estimator," *Journal of arXiv preprint arXiv:1011.1828*, 2010.

[56] M. He, J. Zhang, and V. Vittal, "A data mining framework for online dynamic security assessment: Decision trees, boosting, and complexity analysis," in *Proccedings of IEEE International Conference on Innovative Smart Grid Technologies (ISGT)*, 2012, pp. 1–8.

[57] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *Journal of IEEE Security & Privacy*, no. 3, pp. 75–77, 2009.

[58] A. Hahn and M. Govindarasu, "Smart grid cybersecurity exposure analysis and evalution framework," in *Proccedings of IEEE International Conference on Power and Energy Society General Meeting*, 2010, pp. 1–6.

[59] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 326–333, 2011.

[60] S. McLaughlin, D. Podkuiko, S. Miadzvezhanka, A. Delozier, and P. McDaniel, "Multi-vendor penetration testing in the advanced metering infrastructure," in *Proceedings of ACM Internatinal Conference on the 26th Annual Computer Security Applications*, 2010, pp. 107–116.

[61] J. Lin, W. Yu, X. Yang, G. Xu, and W. Zhao, "On false data injection attacks against distributed energy routing in smart grid," in *Proccedings of IEEE International Conference on Cyber-Physical Systems (ICCPS)*, 2012, pp. 183–192.

[62] W.-T. Tsai, X. Sun, and J. Balasooriya, "Service-oriented cloud computing architecture," in *IEEE Transactions on Information Technology: New Generations (ITNG)*, 2010, pp. 684–689.

[63] Overview: Nist cloud computing efforts, nist senior executive for cloud computing. [Online]. Available: http://csrc.nist.gov/groups/SNS/cloud-computing/documents/forumworkshop-may2010/nist_cloud_computing_forum-leaf.pdf

[64] C. Vecchiola, X. Chu, and R. Buyya, "Aneka: a software platform for .net-based cloud computing," *Journal of High Speed and Large Scale Scientific Computing*, vol. 18, pp. 267–295, 2009.

[65] Y. Huang, H. Su, W. Sun, J. M. Zhang, C. J. Guo, J. M. Xu, Z. B. Jiang, S. X. Yang, and J. Zhu, "Framework for building a low-cost, scalable, and secured platform for web-delivered business services," *IBM Journal of Research and Development*, vol. 54, no. 6, pp. 4–1, 2010.

[66] W. Zeng, Y. Zhao, K. Ou, and W. Song, "Research on cloud storage architecture and key technologies," in *Proceedings of the 2nd ACM International Conference on Interaction Sciences: Information Technology, Culture and Human*, 2009.

[67] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "Racs: a case for cloud storage diversity," in *Proceedings of the 1st ACM symposium on Cloud Computing*, 2010.

[68] S. Kamara and K. Lauter, "Cryptographic cloud storage," *Journal of Financial Cryptography and Data Security*, pp. 136–149, 2010.

[69] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage," *Journal of IEEE Security & Privacy*, vol. 8, no. 6, pp. 40–47, 2010.

[70] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Proccedings of IEEE International Conference on Grid Computing Environments Workshop (GCE)*, 2008.

[71] G. Wang, Q. Liu, and J. Wu, "Hierarchical attribute-based encryption for fine-grained access control in cloud storage services," in *Proceedings of the 17th ACM conference on Computer and communications security*, 2010.

[72] P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin, and M. Walfish, "Depot: Cloud storage with minimal trust," *ACM Transactions on Computer Systems (TOCS)*, vol. 29, no. 4, p. 12, 2011.

[73] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," in *Proceedings of the 6th IEEE Internatinal Conference on Symposium on Opearting Systems Design & Implementation - Volume 6*, 2004.

[74] J. Lin and C. Dyer, *Data-intensive Text Processing with MapReduce*, ser. G - Reference, Information and Interdisciplinary Subjects Series. Morgan & Claypool, 2010.

[75] J. T. Morken, "Distributed netflow processing using the map-reduce model," *PHD Thesis, Norwegian University of Science and Technology*, 2010.

[76] M. Ebrahimi, "Solving linear programs in mapreduce," *Master Thesis, Universität des Saarlandes*, 2011.

[77] D. Alves, P. Bizarro, and P. Marques, "Flood: Elastic streaming mapreduce," in *Proceedings of the 4th ACM International Conference on Distributed Event-Based Systems*, 2010.

[78] C. Doulkeridis and K. Norvag, "On saying enough already! in mapreduce," in *Proceedings of the ACM 1st International Workshop on Cloud Intelligence*, 2012.

[79] F. Halim, R. H. Yap, and Y. Wu, "A mapreduce-based maximum-flow algorithm for large small-world network graphs," in *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2011.

[80] S. Ostermann, A. Iosup, N. Yigitbasi, R. Prodan, T. Fahringer, and D. Epema, "A performance analysis of ec2 cloud computing services for scientific computing," *Journal of Cloud Computing*, pp. 115–131, 2010.

[81] T. Gunarathne, T.-L. Wu, J. Qiu, and G. Fox, "Mapreduce in the clouds for science," in *Proceedings of IEEE 2nd International Conference on Cloud Computing Technology and Science (CloudCom)*, 2010.

[82] Hyrax: cloud computing on mobile devices using mapreduce.

[83] H. Shivhare, N. Mishra, and S. Sharma, "Cloud computing and big data," in *Proceedings of IEEE International Conference on Cloud, Big Data and Trust*, 2013.

[84] Z. Chen, F. Han, J. Cao, X. Jiang, and S. Chen, "Cloud computing-based forensic analysis for collaborative network security management system," *Journal of Tsinghua Science and Technology*, vol. 18, no. 1, pp. 40–50, 2013.

[85] H. Liu and D. Orban, "Cloud mapreduce: a mapreduce implementation on top of a cloud operating system," in *Proceedings of the 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 2011.

[86] J. Tan, X. Pan, E. Marinelli, S. Kavulya, R. Gandhi, and P. Narasimhan, "Kahuna: Problem diagnosis for mapreduce-based cloud computing environments," in *Proccedings of IEEE International Conference on Network Operations and Management Symposium (NOMS)*, 2010.

[87] J. Zhang, D. Xiang, T. Li, and Y. Pan, "M2m: A simple matlab-to-mapreduce translator for cloud computing," *Journal of Tsinghua Science and Technology*, vol. 18, no. 1, pp. 1–9, 2013.

[88] C. Lam, *Hadoop in action*. Manning Publications Co., 2010.

[89] T. White, *Hadoop: the definitive guide*. O'Reilly, 2012.

[90] D. Borthakur, "The hadoop distributed file system: Architecture and design," *Journal of Hadoop Project Website*, vol. 11, no. 2007, p. 21, 2007.

[91] D. J. Suryawanshi and M. U. Mande, "Traffic measurement and analysis with hadoop," *International Journal of Engineering*, vol. 2, no. 10, 2013.

[92] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proccedings of IEEE International Symposium on Mass Storage Systems and Technologies (MSST)*, 2010.

[93] K. Kambatla, A. Pathak, and H. Pucha, "Towards optimizing hadoop provisioning in the cloud," in *Proceedings of the First Workshop on Hot Topics in Cloud Computing*, 2009.

[94] J. Xie, S. Yin, X. Ruan, Z. Ding, Y. Tian, J. Majors, A. Manzanares, and X. Qin, "Improving mapreduce performance through data placement in heterogeneous hadoop clusters," in *Proccedings of IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW)*, 2010.

[95] T. Sandholm and K. Lai, "Dynamic proportional share scheduling in hadoop," *Journal of Job Scheduling Strategies For Parallel Processing*, pp. 110–131, 2010.

[96] J. Shafer, S. Rixner, and A. L. Cox, "The hadoop distributed filesystem: Balancing portability and performance," in *Proceedings of IEEE International Symposium on Performance Analysis of Systems & Software (ISPASS)*, 2010.

[97] Transforming the electricity system to meet future demand and reduce greenhouse gas emissions. [Online]. Available: https://www.cisco.com/web/about/ac79/docs/wp/Smart_Grid_WP_1124a FINAL.pdf

[98] India's massive power blackout: Could smart grids help? [Online]. Available: http://www.nature.com/ncomms/journal/v2/n11/abs/ncomms1563.html?WT.ec_idN̄COMMS-20111122

[99] Is this the world's biggest power cut? chaos strikes for a second day running as over 600 million people endure blackout in india. [Online]. Available: http://www.dailymail.co.uk/news/article-2181517/Indias-power-grid-fails-second-day-running-600-million-people-endure-blackout.html

[100] S. Oren, D. Callaway, T. Mount, M. Zhang, R. Thomas, G. Gross, and A. Dominguez-Garcia., *Renewable Energy Integration and the Impact of Carbon Regulation on the Electric Grid*. Power Systems Engineering Research Center (PSERC), 2012.

[101] Daily renewables watch. [Online]. Available: http://www.caiso.com/market/Pages/ReportsBulletins/DailyRenewablesWatch.aspx

[102] 2007 long-term reliability assessment 20072016. [Online]. Available: http://www.nerc.com/files/LTRA2007.pdf

[103] *Fourth Assessment Report: Climate Change*. [Online]. Available: https://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr_full_report.pdf

[104] The economics of climate change: Executive summary. [Online]. Available: http://siteresources.worldbank.org/INTINDONESIA/Resources/226271-1170911056314/3428109-1174614780539/SternReviewEng.pdf

[105] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Journal of Biometrika*, vol. 52, no. 3-4, 1965.

[106] M. B. Wilk and R. Gnanadesikan, "Probability plotting methods for the analysis of data," *Journal of Biometrika*, vol. 55, no. 1, 1968.

[107] How to calculate transmission line losses.

[108] Copper hexacyanoferrate battery electrodes with long cycle life and high power. [Online]. Available: http://www.nature.com/ncomms/journal/v2/n11/abs/ncomms1563.html?WT.ec_id=NCOMMS-20111122

[109] Electric power monthly. http://www.eia.gov/electricity/monthly/pdf/epm.pdf.

[110] J. Shi, J. Wan, H. Yan, and H. Suo, "A survey of cyber-physical systems," in *Proccedings of IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, 2011, pp. 1–6.

[111] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems." in *Proccedings of IEEE International Conference on HotSec*, 2008.

[112] Guidelines for smart grid cyber security. [Online]. Available: http://csrc.nist.gov/publications/PubsNISTIRs.html#NIST-IR-7628

[113] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in *Proccedings of IEEE International Conference on Critical Information Infrastructures Security*, 2010, pp. 176–187.

[114] F. M. Cleveland, "Cyber security issues for advanced metering infrasttructure (ami)," in *Proccedings of IEEE International Conference on Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–5.

[115] K. Song, D. Seo, H. Park, H. Lee, and A. Perrig, "Omap: One-way memory attestation protocol for smart meters," in *Proccedings of IEEE International Symposium on Parallel and Distributed Processing with Applications Workshops (ISPAW)*, 2011, pp. 111–118.

[116] W. Yu, “False data injection attacks in smart grid: Challenges and solutions,” in *Proceeding of NIST Cyber Security for Cyber-Physical System (CPS) Workshop*, 2012.

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

[118] J. Minkel, “The 2003 northeast blackout–five years later,” *Journal of Scientific American*, vol. 13, 2008.

[119] J. Vijayan, “Stuxnet renews power grid security concerns,” *Journal of Computerworld*, vol. 26, 2010.

[120] Smart meter security: a survey. [Online]. Available: http://www.cl.cam.ac.uk/řja14/Papers/JSAC-draft.pdf

[121] Q. Yang, D. An, and W. Yu, “On time desynchronization attack against ieee 1588 protocol in power grid systems,” in *Proccedings of IEEE International Conference on Energytech*, 2013, pp. 1–5.

[122] B. Zhu, A. Joseph, and S. Sastry, “A taxonomy of cyber attacks on scada systems,” in *Proccedings of IEEE International Conference on Internet of things (iThings/CPSCom), and 4th international conference on cyber, physical and social computing*, 2011, pp. 380–388.

[123] P. S. M. Pires and L. A. H. Oliveira, “Security aspects of scada and corporate network interconnection: An overview,” in *Proccedings of IEEE International Conference on Dependability of Computer Systems*, 2006, pp. 127–134.

[124] P. A. Ralston, J. H. Graham, and J. L. Hieb, “Cyber security risk assessment for scada and dcs networks,” *IEEE Transaction on ISA*, vol. 46, no. 4, pp. 583–594, 2007.

[125] S. Hong and M. Lee, "Challenges and direction toward secure communication in the scada system," in *Proccedings of IEEE International Conference on Communication Networks and Services Research Conference (CNSR)*, 2010, pp. 381–386.

[126] S. Karnouskos, "Stuxnet worm impact on industrial cyber-physical system security," in *Proccedings of IEEE International Conference on Industrial Electronics Society*, 2011, pp. 4490–4494.

[127] H. Li, L. Lai, and W. Zhang, "Communication requirement for reliable and secure state estimation and control in smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp. 476–486, 2011.

[128] L. Sankar, S. Kar, R. Tandon, and H. V. Poor, "Competitive privacy in the smart grid: An information-theoretic approach," in *Proccedings of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 220–225.

[129] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, K. Poolla *et al.*, "Smart grid data integrity attacks: characterizations and countermeasures π," in *Proccedings of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 232–237.

[130] F. Pasqualetti, R. Carli, and F. Bullo, "A distributed method for state estimation and false data detection in power networks," in *Proccedings of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 469–474.

[131] O. Vuković, K. C. Sou, G. Dán, and H. Sandberg, "Network-layer protection schemes against stealth attacks on state estimators in power systems," in *Proccedings of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 184–189.

[132] Y. Liu, P. Ning, H. Dai, and A. Liu, "Randomized differential dsss: Jamming-resistant wireless broadcast communication," in *Proccedings of IEEE International Conference on INFOCOM*, 2010, pp. 1–9.

[133] D. Jin, D. M. Nicol, and G. Yan, "An event buffer flooding attack in dnp3 controlled scada systems," in *Proccedings of IEEE International Conference on Winter Simulation*, 2011, pp. 2619–2631.

[134] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proccedings of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 214–219.

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

[136] State electricity profiles 2009. [Online]. Available: http://www.eia.gov/

[137] A. B. Nagarajan, F. Mueller, C. Engelmann, and S. L. Scott, "Proactive fault tolerance for hpc with xen virtualization," in *Proceedings of the 21st annual international conference on Supercomputing*. ACM, 2007, pp. 23–32.

[138] A.-M. Juuso, A. Takanen, and K. Kittilä, "Proactive cyber defense: Understanding and testing for advanced persistent threats (apts)," in *Proceedings of the 12th European Conference on Information Warfare and Security: ECIW 2013*, 2013, p. 383.

[139] Sysstat utilities home page. [Online]. Available: http://sebastien.godard.pagesperso-orange.fr

[140] M. Cramer, J. Cannady, and J. Harrell, "New methods of intrusion detection using control-loop measurement," in *Proceedings of the Technology in Information Security Conference (TISC)*, vol. 95, 1995, pp. 1–10.

[141] Y. Wang, H. Yang, X. Wang, and R. Zhang, "Distributed intrusion detection system based on data fusion method," in *Proccedings of IEEE International Conference on Intelligent Control and Automation*, vol. 5, 2004, pp. 4331–4334.

[142] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Journal of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[143] N. Zhang, W. Yu, X. Fu, and S. K. Das, "Maintaining defender's reputation in anomaly detection against insider attacks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 597–611, 2010.

[144] W. Yu and K. Liu, "Game theoretic analysis of cooperation stimulation and security in autonomous mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 5, pp. 507–521, 2007.

[145] T. Alpcan and T. Başar, "A game theoretic analysis of intrusion detection in access control systems," in *IEEE Conference on Decision and Control*, vol. 2, 2004, pp. 1568–1573.

[146] Y. Liu, C. Comaniciu, and H. Man, "A bayesian game approach for intrusion detection in wireless ad hoc networks," in *Proccedings of ACM International Workshop on Game Theory for Communications and Networks*, 2006, p. 4.

[147] A. B. Sharma, F. Ivančić, A. Niculescu-Mizil, H. Chen, and G. Jiang, "Modeling and analytics for cyber-physical systems in the age of big data," *Journal of ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 4, pp. 74–77, 2014.

[148] L.-A. Tang, J. Han, and G. Jiang, "Mining sensor data in cyber-physical systems," *Journal of Tsinghua Science and Technology*, vol. 19, no. 3, pp. 225–234, 2014.

[149] C.-H. Chen, C.-Y. Chen, C.-H. Hsia, and G.-X. Wu, “Big data collection gateway for vision-based smart meter reading network,” in *Procceddings of IEEE International Congress on Big Data (BigData Congress)*, 2014, pp. 266–269.

[150] F. Li, B. C. Ooi, M. T. Özsu, and S. Wu, “Distributed data management using mapreduce,” *Journal of ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 31, 2014.

[151] L. Qin, J. X. Yu, L. Chang, H. Cheng, C. Zhang, and X. Lin, “Scalable big graph processing in mapreduce,” in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. ACM, 2014, pp. 827–838.

[152] W. Yu, H. Zhang, L. Ge, and R. Hardy, “On behavior-based detection of malware on android platform,” in *Proceedings of IEEE Globecom*, 2013.

[153] Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence. [Online]. Available: http://www.stanford.edu/ shoude/FieldExperimentPowermeter_vfinal_July 2011.pdf

Curriculum Vita

Name: Guobin Xu

Research Interests: Cyber-physical System (CPS), Computer Network and Security, Cloud Computing, and Big Data, including smart grid modeling and security, network threat monitoring and detection, and cloud computing based big data processing for network and security management.

Education:

- D.Sc. in Information Technology — September 2011 - Present
  Towson University, Towson, MD, U.S.A.
- M.S. in Applied Information Technology — June 2011
  Towson University, Towson, MD, U.S.A.
- B.S. in Mathematics — June 2009
  Qingdao University, Qingdao, P.R.China.
- B.S. in Economics — June 2009
  Qingdao University, Qingdao, P.R.China.

Teaching Experience:

- Lecturer — 09/2013 - Present
  Department of Computer and Information Sciences, Towson University.
- Teaching Assistant — 09/2011 - 05/2013
  Department of Computer and Information Sciences, Towson University.
- Math Teacher Internship — 01/2008 - 05/2008
  No. 19 high school of Qingdao, Qingdao(P.R.China).

Research Experience:

- Researching Assistant — 01/2011 - Present
  Department of Computer and Information Sciences, Towson University.

Working Experience:

- Graduate Assistant 01/2011 - 06/2011
  Department of Computer and Information Sciences, Towson University.
- Math Lab Tutor 09/2010 - 06/2011
  Department of Mathematics, Towson University.
- Volunteer 07/2008 - 09/2008
  The Olympic organizing committee, Qingdao.
- Tour Guide 07/2007 - 09/2007
  China International Travel Service Corp (CITS) of Qingdao, Qingdao.
- Event and Conference Coordinator 07/2006 - 09/2006
  China International Travel Service Corp (CITS) of Qingdao, Qingdao.

Refereed Book Chapters:

- Wei Yu, **Guobin Xu**, Khanh D. Pham, Erik P. Blasch, Genshe Chen, Dan Shen, and Paul Moulema, "A Framework for Cyber-Physical System Security Situation Awareness", accepted to appear in *Foundational Methods for Cyber-physical Systems*, 2015.
- Jie Lin, Wei Yu, Xinyu yang, and **Guobin Xu**, "Cyber Security in Cyber-Physical Systems: On False Data Injection Attacks in the Smart Grid", accepted to appear in *Cyber-Physical System Design with Sensor Networking Technologies*, published by IET (formerly IEE) Press in London, England, 2015.
- Linqiang Ge, Hanling Zhang, **Guobin Xu**, Wei Yu, Chen Chen, and Erik P. Blasch, "Towards MapReduce Based Machine Learning Techniques for Processing Massive Network Threat Monitoring Data", accepted to appear in *Networking for Big Data*, published by CRC Press & Francis Group, USA, 2015.

- Wei Yu, Linqiang Ge, **Guobin Xu**, and Xinwen Fu, “Towards Neural Network Based Malware Detection On Android Mobile Devices", accepted to appear in *Springer Book Series: Cybersecurity Systems for Human Cognition Augmentation*, 2014.

Refereed Journal Publications:

- **Guobin Xu**, Wei Yu, Zhijiang Chen, Hanlin Zhang, Paul Moulema, Xinwen Fu, and Chao Lu, “A Cloud Computing Based System for Network Security Management", accepted to appear in *International Journal of Parallel, Emergent and Distributed Systems (IJPEDS) - Taylor & Francis*, 30(1):29-45, 2015.

- Wei Yu, Dou An, David Griffith, Qingyu Yang, and **Guobin Xu**, “On Statistical Modeling and Forecasting of Energy Usage in Smart Grid", accepted to appear in *the ACM International Journal of Applied Computing Review (ACR)* , 2015.

- Paul Moulema, Wei Yu, **Guobin Xu**, David Griffith, Nada Golmie, Chao Lu, and David Su, “On Effectiveness of Mesh-based Protocols for Smart Grid Communication Networks", accepted to appear in *ACM International Journal of Applied Computing Review (ACR)*, 14(2): 59-70, 2014.

- Jie Lin, Wei Yu, David Griffith, Xinyu Yang, **Guobin Xu**, and Chao Lu, “On Distributed Energy Routing Protocols in the Smart Grid", accepted to appear in *Springer’s Studies in Computational Science* (as one of the top papers selected from the 14th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)), 2013.

Refereed Conference Publications:

- Wei Yu, Dou An, David Griffith, Qingyu Yang, and **Guobin Xu**, “On Statistical Modeling and Forecasting of Energy Usage in Smart Grid", in *Proceedings of ACM International Conference on Reliable & Convergent Systems (RACS)*, October 2014.

- Xin Chen, Wei Yu, David Griffith, Nada Golmie, and **Guobin Xu**, “On Effectiveness of Energy Storage System against Smart Grid Cascading Failure", in *Proceedings of ACM International Conference on Reliable & Convergent Systems (RACS)*, October 2014.

- Nnanna Ekedebe, Zhijiang Chen, **Guobin Xu**, Chao Lu, and Wei Yu, “On an Efficient and Effective Intelligent Transportation System (ITS) Using Field and Simulation Data", in *Proceedings of SPIE Defense, Security, and Sensing (DSS)*, May 2014.

- Wei Yu, Zhijiang Chen, **Guobin Xu**, Sixiao Wei, and Nnanna Ekedebe, “A Threat Monitoring System in Enterprise Networks with Mobile Devices", in *Proceedings of ACM International Conference on Reliable & Convergent Systems (RACS)*, October 2013.

- Paul Moulema, Wei Yu, **Guobin Xu**, David Griffith, Nada Golmie, Chao Lu, and David Su, “On Simulation Study of Mesh-based Protocols for Smart Grid Communication Networks", in *Proceedings of ACM International Conference on Reliable & Convergent Systems (RACS)*, October 2013.

- Wei Yu, **Guobin Xu**, Zhijiang Chen, and Paul Moulema, “A Cloud Computing Based Architecture for Cyber Security Situation Awareness", in *Proceedings of the 4th International Workshop on Security and Privacy in Cloud Computing (SPCC)*, October 2013.

- **Guobin Xu**, Paul Moulema, and Wei Yu, “Integrating Distributed Energy Resources in Smart Grid: Modeling and Analysis", in *Proceedings of IEEE EnergyTech 2013*, May 2013.

- Wei Yu, Sixiao Wei, **Guobin Xu**, Genshe Chen, Khanh Pham, Erik P. Blasch, and Chao Lu, “On Effectiveness of Routing Algorithms for Satellite Communication Networks", in *Proceedings of SPIE Defense, Security, and Sensing 2013*,

April/May 2013.

- Wei Yu, Hanlin Zhang, and **Guobin Xu**, “A Study of Malware Detection on Smart Mobile Devices", in *Proceedings of SPIE Defense, Security, and Sensing 2013*, April/May 2013.

- Jie Lin, Wei Yu, Xinyu Yang, **Guobin Xu**, and Wei Zhao, “On False Data Injection Attacks against Distributed Energy Routing in Smart Grid", in *Proceedings of the 3rd ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, April 2012

Honors and Awards:

- Graduate Student Association Award 2013
  Towson University

- Travel Grant Awards 2013
  The 4th International Workshop on Security and Privacy in Cloud Computing (CNS-SPCC)

- University Scholarship Awards 2005 - 2009
  Qingdao University