Energy Systems

Vijay Pappu
Marco Carvalho
Panos M. Pardalos *Editors*

Optimization and Security Challenges in Smart Power Grids

Springer

Energy Systems

Series Editor

Panos M. Pardalos, Gainesville FL, USA

For further volumes:
http://www.springer.com/series/8368

Vijay Pappu · Marco Carvalho
Panos M. Pardalos
Editors

Optimization and Security Challenges in Smart Power Grids

Springer

Editors
Vijay Pappu
Panos M. Pardalos
Industrial and Systems Engineering
University of Florida
Gainesville, FL
USA

Marco Carvalho
Florida Institute of Technology
Melbourne, FL
USA

ISSN 1867-8998 ISSN 1867-9005 (electronic)
ISBN 978-3-642-38133-1 ISBN 978-3-642-38134-8 (eBook)
DOI 10.1007/978-3-642-38134-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013947566

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The electrical power grid is often referred to as one of the most complex man-made systems on Earth. Its importance to all aspects of our daily lives, economic stability, and national security cannot be overstated, and the need for an updated, secure, resilient, and smarter power grid infrastructure is increasingly recognized and supported by policy makers and market forces.

The promise of a smarter electrical grid is likely to be one the most important transformational changes in our national power infrastructure. This could significantly affect how consumers use and pay for their electrical power, thus fundamentally changing the power industry that we know today.

Smart Grid technologies combine power generation and delivery systems with advanced communication systems to help save energy, reduce energy costs, and improve reliability. The combination of these technologies enable new approaches for load balancing and power distribution, allowing for optimal runtime power routing, and cost management. Such unprecedented capabilities, however, also introduce new sets of challenges at the technical and regulatory levels that must be addressed by the industry and the research community. This book, organized as a part of the workshop 'Systems and Optimization Aspects of Smart Grid Challenges' brings together a number of perspectives and approaches to smart grid challenges and optimization.

This book primarily covers both the optimization and the security aspects of smart grid technologies. From a control and optimization perspective, the book includes chapters on unit commitment, homeostatic control, flexible demands, and others. From a cyber security perspective, the book incudes chapters on secure sensor measurements, temper detection, and proposed approaches to trustworthy architectures, among others. These articles address some of the many important aspects in smart grids control and optimization research.

We would like to express our gratitude to all the reviewers and contributing authors for offering their expertise and providing valuable material used to compose this volume. We thank Springer for the opportunity to make a contribution in advancing and sharing the state-of-the-art research in smart grid technologies.

Vijay Pappu
Marco Carvalho
Panos M. Pardalos

Contents

Optimization Approaches to Security-Constrained Unit Commitment and Economic Dispatch with Uncertainty Analysis 1
Dzung T. Phan and Ali Koc

Homeostatic Control and the Smart Grid: Applying Lessons from Biology 39
Martin Beckerman

Operator's Interruption-Cost-Based Sectionalization Method for 3-Feeder Radial Distribution Architecture 53
Virginijus Radziukynas, Neringa Radziukynienė, Arturas Klementavičius and Darius Naujokaitis

The Role of Flexible Demands in Smart Energy Systems 79
Kristin Dietrich, Jesus M. Latorre, Luis Olmos and Andres Ramos

Smart Grid Tamper Detection Using Learned Event Patterns 99
William L. Sousan, Qiuming Zhu, Robin Gandhi and William Mahoney

Automating Electric Substations Using IEC 61850 117
Peter J. Hawrylak, Jeyasingam Nivethan and Mauricio Papa

Phasor Measurement Unit and Phasor Data Concentrator Cyber Security 141
Thomas H. Morris, Shengyi Pan, Uttam Adhikari, Nicolas Younan, Roger King and Vahid Madani

Infrastructure Security for Smart Electric Grids: A Survey 161
Naran M. Pindoriya, Dipankar Dasgupta, Dipti Srinivasan and Marco Carvalho

Known Secure Sensor Measurements Concept and Its Application for Critical Infrastructure Systems 181
Annarita Giani, Ondrej Linda, Milos Manic and Miles McQueen

Data Diodes in Support of Trustworthy Cyber Infrastructure and Net-Centric Cyber Decision Support 203
H. Okhravi, F. T. Sheldon and J. Haines

Index 217

Optimization Approaches to Security-Constrained Unit Commitment and Economic Dispatch with Uncertainty Analysis

Dzung T. Phan and Ali Koc

Abstract At the heart of the future smart grid lie two related challenging optimization problems: unit commitment and economic dispatch. The contemporary practices such as intermittent renewable power, distributed generation, demand response, etc., induce uncertainty into the daily operation of an electric power system, and exacerbate the ability to handle the already complicated intermingled problems. We introduce the mathematical formulations for the two problems, present the current practice, and survey solution methods for solving these problems. We also discuss a number of important avenues of research that will receive noteworthy attention in the coming decade.

Keywords Economic dispatch · Power flow · Uncertainty · Stochastic · Security-constrained · Unit commitment

1 Introduction

At the heart of the future smart grid lie two related challenging optimization problems: unit commitment (UC) and economic dispatch (ED). When operational and physical constraints are considered not only under normal operating conditions, but also under contingency conditions, the UC and ED problem becomes the security-constrained UC and ED problem. We focus on UC and ED in this chapter because these two problems are most relevant to independent system operators (ISOs) and regional transmission organizations (RTOs) daily operation as they need to be solved on a

D. T. Phan (✉) · A. Koc
Business Analytics and Mathematical Sciences Department,
IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA
e-mail: phandu@us.ibm.com

A. Koc
e-mail: akoc@us.ibm.com

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_1,

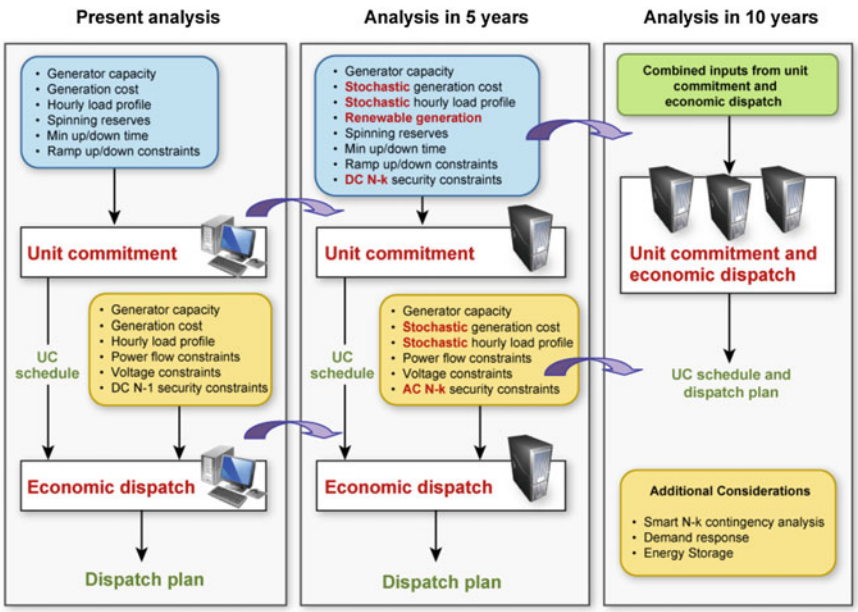

Fig. 1 Present and future of UC and ED

daily basis, which require both computational and algorithmic improvements to meet real-time operational requirements.

Although these two problems are intermingled with each other, most of the current theoretical and practical effort treats them separately, because of the computational difficulty of solving a single unified problem. As Fig. 1 illustrates, present solutions to the unit commitment problem consider only a direct current (DC) approximation of the alternate current (AC) transmission constraints. This problem observes any generator-related constraints, demand constraints, and linear transmission constraints. The output of this is an optimal schedule for generators in a twenty-four-hour time horizon, and is given as input to the economic dispatch problem. Economic dispatch problem then handles the original AC power flow constraints and outputs a dispatch plan: how much power to produce from each generator, and how to transmit the power over the network. To account for unexpected failure of generators and transmission lines, current unit commitment practices enforce spinning reserve requirements, allocating a fraction of a generator's capacity to reserves. Similar contingency analysis is also performed in economic dispatch, making sure that the load at each node of the network can be satisfied in the case of a failure of one of the generators, transmission lines, or other devices, which is also called N-1 contingency analysis.

There are several points that current practice is missing and that need to be handled in the very near future: integration of renewable energy into the grid, considering failure of more than one generator and/or transmission line, also called N-k

contingency analysis, considering stochasticities in the problem, such as generation costs and load profiles, and being able to solve larger instances of both unit commitment and economic dispatch problems. Theoretical and practical efforts along these lines have gained momentum and will likely keep increasing (Acar et al. 2011).

The ultimate goal is to solve practical instances of these two problems together, considering other relevant issues such as "demand response" and "energy storage." As will be discussed, this definitely requires more algorithmic advancement that uses high-performance and parallel computing environments. Today, a typical commercial integer programming solver can handle a unit commitment problem with 100 units, 24 time periods, and 50 uncertainty scenarios. Real-life instances consist of several thousand buses, more than one thousand generators, 48–72 time periods, more than one hundred contingencies, and a few hundreds of scenarios. Solving such a large-scale real-life instance of the unit commitment and economic dispatch problem together, along with all other relevant issues, is a grand challenge that will reinforce need for fast and parallelizable decomposition algorithms.

Both ED and UC can be formulated as nonlinear optimization problems (NLP) and mixed integer NLPs that are, in general, non-convex and nonlinear. Existing industrial solutions to these two problems have been traditionally dominated by the Lagrangian relaxation methods, and only recently have been using general-purpose integer programming solvers. Academic solutions are more diverse, but they are typically demonstrated on much smaller IEEE bus cases than the real-life scenarios.

Existing solutions have a number of limitations; failure to solve real-life problem instances; the sub-optimality of the solutions; inability to guarantee the convergence to a feasible solution; insufficient treatment of contingency scenarios—typically focusing on N-1 but not on N-k contingencies; limited consideration of the uncertainty existing in various forms such as in loading and generation, transmission outages, fuel prices, renewables, etc. These limitations require current power systems operate under very conservative standards and maintain excessive margins in order to address all types of unmodeled uncertainties. These excessive margins significantly limit the efficiency of the power grid.

The formulations of both unit commitment and economic dispatch for system operation needs to be improved in order to obtain significant integration of intermittent renewable energy generation and enable demand response. Advanced optimization techniques targeting globally optimal solutions and with guaranteed convergence need to be developed. Integrating renewables and addressing N-k contingencies is required to support real-time secure operations. Solutions to both unit commitment and economic dispatch problems need to be implemented in a hybrid computing environment that supports: evaluation of multiple scenarios resulting from contingencies and loading/generation profiles in parallel; and parallel execution of decomposition algorithms for large-scale optimization problems with guaranteed convergence to high-quality solutions.

In the next section, we give a mathematical formulation for the unit commitment and economic dispatch problems, discussing alternative system constraints and objective functions. In Sect. 3, we start with the unit commitment problem with linearized power flow constraints, and discuss several exact and heuristic algorithms

developed in the literature. In Sect. 4, we concentrate on the uncertainty affecting the unit commitment problem, and summarize the approaches to handle the uncertainty. In Sect. 5, we continue with the economic dispatch problem with AC power flow constraints, and discuss various solution methodologies suggested in the literature. In Sect. 6, we discuss the literature on the security-constrained economic dispatch problem. Finally, in the last section, we conclude and point out the future research directions.

2 Overview of Unit Commitment and Economic Dispatch Formulation

Several variations of UC and ED problem have been studied in the literature. We present an overview of objective functions and constraints in a unified framework, and then we point out the constraints and objective functions we consider in each individual problem. An electric grid management entity controls the dispatching of generation units over a network of multiple local buses interconnected via transmission lines. We let $\mathscr{B}$ denote the set of buses (nodes) in the grid network indexed by b, $\mathscr{L}$ denote the set of transformers indexed by l, $\mathscr{T}$ denote the set of time periods indexed by t, $\mathscr{I}$ denote the set of generators (units) indexed by i, $\mathscr{I}_b$ denote the set of units connected to bus $b \in \mathscr{B}$, and $\mathscr{N}_b$ denote the set of adjacent buses of b. Power transmission over a network of transmission lines that connect the buses is determined by the voltage sets at the buses. We let $\mathbf{V}_t$ denote the vector of voltages, whose entries are $V_{bt}, \quad b \in \mathscr{B}, t \in \mathscr{T}$. We note that throughout the text we use the terms "node" and "bus," and the terms "unit" and "generator" interchangeably, preferring to use the term "unit" mostly in the context of the UC problem and the term "generator" in the context of the ED problem.

2.1 Constraints of the Problem

2.1.1 Power Flow Constraints

Current flow between any two connected nodes b and b' is determined by the voltage difference. At time period $t \in \mathscr{T}$, the complex nodal current injections $\mathbf{I}^t$ can be computed by the bus admittance matrix $\mathbf{Y}^t$ and the complex nodal voltages $\mathbf{V}_t$:

$$\mathbf{I}^t = \mathbf{Y}^t \cdot \mathbf{V}_t.$$

Net nodal power injections are then expressed as $\mathbf{V}_t \cdot \mathbf{I}^{t*}$, where $\mathbf{I}^{t*}$ is the complex conjugate of $\mathbf{I}^t$. Therefore, the active power PF^t and reactive power QF^t of the nodal the complex power are functions of voltages $\mathbf{V}_t$:

$$\mathbf{V}_t \cdot \mathbf{I}^{t*} = PF^t(\mathbf{V}_t) + \mathrm{j}QF^t(\mathbf{V}_t),$$

where j is the imaginary unit. The complex nodal voltage can be written in either rectangular form $V_{bt} = e_{bt} + \mathrm{j}f_{bt}$, or in polar form $V_{bt} = |V_{bt}|(\cos\theta_{bt} + \mathrm{j}\sin\theta_{bt})$. Both power balance functions, $PF_b^t(\mathbf{V}_t)$ and $QF_b^t(\mathbf{V}_t)$, are quadratic in the real and imaginary parts of the base transmission variables $\mathbf{V}_t$, and is non-convex. The AC power flow constraints for bus b at time period t reads:

$$PF_b^t(\mathbf{V}_t) = \sum_{i \in \mathscr{I}_b} p_{it} - p_{bt}^d, b \in \mathscr{B}, \quad t \in \mathscr{T}, \tag{1a}$$

$$QF_b^t(\mathbf{V}_t) = \sum_{i \in \mathscr{I}_b} q_{it} - q_{bt}^d, b \in \mathscr{B}, \quad t \in \mathscr{T}, \tag{1b}$$

where p_{bt}^d and q_{bt}^d are the active and reactive power demands (loads) at node b at time period t, and p_{it} and q_{it} are the active and reactive power generation at unit i at time period t.

When considered along with a UC problem, these nonlinear AC power flow constraints complicate the already complicated integer programming problem, and is usually impractical to handle. It is a common approach to use a linearized DC approximation, i.e., DC power flows, where reactive power equations are ignored, and the magnitude of voltages are assumed to be a unit, only phase angles and active power balances are considered:

$$\widehat{PF}_b^t(\boldsymbol{\theta}_t) = \sum_{b' \in \mathcal{N}_b} \frac{\theta_{bt} - \theta_{b't}}{x_{bb'}^t} = \sum_{i \in \mathscr{I}_b} p_{it} - p_{bt}^d, \quad b \in \mathscr{B}, \quad t \in \mathscr{T}, \tag{2}$$

where $x_{bb'}^t$ is the reactance between nodes b and b' at time period t.

2.1.2 Ramping Constraints

Ramping constraints restrict the amount of increase and decrease in the active power offer of a unit:

$$p_{it} - p_{i,t-1} \leq \overline{R}_i, i \in \mathscr{I}, \ t \in \mathscr{T}, \tag{3a}$$

$$p_{i,t-1} - p_{it} \leq \underline{R}_i, i \in \mathscr{I}, \ t \in \mathscr{T}, \tag{3b}$$

where $\overline{R}_i$ and $\underline{R}_i$ denote the ramping up and down limit of unit i.

Constraints (3a) and (3b) model the general case of ramping limits, not differentiating between the cases of a unit being up or down in any of the two consecutive time periods. Some thermal units, however, have startup or shutdown ramping limits that are different than those when unit is up in both of the consecutive time periods. Some units, even, cannot be ramped up to its minimum generation level, or, similarly,

some cannot be ramped down to zero in a single time period. Arroyo and Conejo (2004) give a detailed ramping model for such units.

2.1.3 Minimum Up and Down Time Constraints

A set of constraints commonly used in the literature is

$$u_{it} - u_{i,t-1} \leq u_{i\tau}, \qquad \tau \in \{t, \ldots, \min\{t + \overline{M}_i - 1, |\mathscr{T}|\}\}, \quad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{4a}$$

$$u_{i,t-1} - u_{it} \leq 1 - u_{i\tau}, \qquad \tau \in \{t, \ldots, \min\{t + \underline{M}_i - 1, |\mathscr{T}|\}\}, \quad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{4b}$$

where $\overline{M}_i$ denote the minimum number of time periods that unit i has to stay up once it is turned on, $\underline{M}_i$ denote the minimum number of time periods that unit i has to stay down once it is turned off, and u_{it} denote the binary up and down status of unit i at time period t.

Rajan and Takriti (2005) formulate the minimum up/down time constraints as

$$\sum_{\substack{\tau=\max\{1,\ t-\overline{M}_i+1\}\\ \min\{|\mathscr{T}|,\ t+\underline{M}_i\}}}^{t} s_{i\tau} \leq u_{it}, \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{5a}$$

$$\sum_{\tau=t+1} s_{i\tau} \leq 1 - u_{it}, \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{5b}$$

$$s_{it} \geq u_{it} - u_{i,t-1}, \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{5c}$$

$$s_{it} \in [0, 1], \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{5d}$$

where a new variable, s_{it}, is introduced to denote whether the unit is started up in the current time period. The authors prove that these four sets of constraints model the minimum up/down time polytope if the cost function does not include a shutdown cost but includes a startup cost that depends only on the current and the previous periods.

2.1.4 Limits on Variables

Physical constraints require upper and lower bound limits on the decision variables. Active and reactive power limits are

$$\underline{P}_i u_{it} \leq p_{it} \leq \overline{P}_i u_{it}, \quad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{6a}$$

$$\underline{Q}_i u_{it} \leq q_{it} \leq \overline{Q}_i u_{it}, \quad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{6b}$$

where $\overline{P}_i$ and $\underline{P}_i$ denote the upper and lower limits on the active power offer of unit i , and similarly $\overline{Q}_i$ and $\underline{Q}_i$ denote the upper and lower limits on the reactive power offer of unit i. Constraints (6a) and (6b) perform two functions: If the unit is down, they force the power to be zero; and, if the unit is up, they arrange the power level to be between the upper and lower limits.

Limits on voltage magnitudes are modeled as

$$\underline{V}_b \leq |V_{bt}| \leq \overline{V}_b, \quad b \in \mathscr{B}, \ t \in \mathscr{T}; \tag{7}$$

limits on voltage angles between nodes are modeled as

$$\underline{\theta}_{bb'} \leq \theta_{bt} - \theta_{b't} \leq \overline{\theta}_{bb'}, \quad b, b' \in \mathscr{B}, \ t \in \mathscr{T}; \tag{8}$$

limits on tap positions of transformers are modeled as

$$\underline{t}_l \leq t_{lt} \leq \overline{t}_l, \quad l \in \mathscr{L}, \ t \in \mathscr{T}; \tag{9}$$

limits on phase shift angles of transformers are modeled as

$$\underline{\phi}_l \leq \phi_{lt} \leq \overline{\phi}_l, \quad l \in \mathscr{L}, \ t \in \mathscr{T}; \tag{10}$$

and the line flow limit are modeled as

$$FL^t_{bb'}(\mathbf{V}_t) \leq \overline{FL}_{bb'}, \quad b, b' \in \mathscr{B}, \quad t \in \mathscr{T}, \tag{11}$$

where $\overline{V}_b$ and $\underline{V}_b$ denote the upper and lower limits on the voltage magnitude at bus b, $\overline{\theta}_{bb'}$ and $\underline{\theta}_{bb'}$ denote the upper and lower limits on the voltage angle between buses b and b', $\overline{t}_l$ and $\underline{t}_l$ denote the upper and lower limits on the tap positions of transformer l, $\overline{\phi}_l$ and $\underline{\phi}_l$ denote the upper and lower limits on the phase shift angle of transformer l, and $\overline{FL}_{bb'}$ denotes the upper limit on the power flow between buses b and b'. The transmission line flow limits (11) can be on apparent power flows $|V_{bt} {I^t_{bb'}}^*|$, active power flows $Real(V_{bt} {I^t_{bb'}}^*)$, or current magnitudes $|I^t_{bb'}|$. Note that the changes in the transformer parameters affect the admittance matrix $\mathbf{Y}^t$. In practice, additional physical and operational constraints can be imposed.

2.2 Objectives of the Problem

The objective function usually includes fuel cost, active power losses, voltage stability, startup cost, and shutdown cost of units:

$$\text{Cost} = \sum_{t \in \mathscr{T}} \left(f^p(\mathbf{p}_t) + f^v(\mathbf{V}_t) + f^q(\mathbf{q}_t) \right) + \sum_{t \in \mathscr{T}, i \in \mathscr{I}} \left(S_i(\overline{\mathbf{u}}_{it}) + H_i(\overline{\mathbf{u}}_{it}) \right), \tag{12}$$

where

- $\mathbf{p}_t = \{p_{it} : i \in \mathscr{I}\}$ and $\mathbf{q}_t = \{q_{it} : i \in \mathscr{I}\}$ are the active and reactive power offers at time period t,
- $\overline{\mathbf{u}}_{it} = (u_{i,1}, \ldots, u_{i,t})$, u_{it} is the binary up and down status of unit i at time period t,
- $f^p(\cdot)$ can be the fuel cost functions of unit i for active power generation or the active power losses over all nodes,
- $f^q(\cdot)$ and $f^v(\cdot)$ can be the active power transmission losses or the deviation from a specified point of control variables, and
- $S_i(\cdot)$ and $H_i(\cdot)$ are startup and shutdown costs of unit i.

A cost-based objective function dominates the studies in the literature, although profit-based terms have started appearing due to the deregulation and restructuring of the power market (Padhy 2004). Fuel cost usually is the sum of quadratic functions of active power p_{it}, i.e., $f^p(\mathbf{p}_t) = \sum_{i\in\mathscr{I}} f^p(p_{it})$, where

$$f^p(p_{it}) = a_i^p\, p_{it}^2 + b_i^p\, p_{it} + c_i^p,$$

although some studies simplify it to linear or piecewise linear functions (Wood and Wollenberg 1996). The constant term is formulated with the help of u_{it} variables as, for instance, $f^p(p_{it}) = a_i^p\, p_{it}^2 + b_i^p\, p_{it} + c_i^p\, u_{it}$. Note that p_{it} is forced to zero when u_{it} is zero by constraint (6a), which forces $f^p(p_{it})$ to be zero when u_{it} is zero.

A possible power loss function can be the sum of active power transmission losses $f^v(\mathbf{V}_t) = \sum_{b,b'\in\mathscr{B}} V_b^t {I_{bb'}^t}^* + \sum V_{b'}^t {I_{b'b}^t}^*$. In tertiary voltage control, the aim is to prevent the voltage drops and losses, the objective is to minimize the weighted sum of the deviations of the control variables $f^q(\mathbf{q}_t) = \sum_{i\in\mathscr{I}} w_i (q_{it} - q_{it}^0)^2$, where w_i are weighting parameters.

It is common to model the startup cost of a unit as a fixed cost that is incurred when the unit is turned on, independent of how long the unit has been offline. Startup cost should be more realistically modeled as a nonlinear (exponential) function of the duration that the unit has been down. Arroyo and Conejo (2000), and Wang and Shahidehpour (1993) approximate the exponential startup cost function by a step function that takes its smallest step value when the unit has been just turned off, and its largest value when the exponential function gets close to its limiting value.

For an ideal unified UC-ED problem, all above-mentioned constraints should be combined into a single optimization problem, and dealt with at once. However, the current optimization methodology advancement hinders this; it is imperative to decompose the unified problem into two parts: the UC and the ED problems, although there exist some attempts to solve the unified problem, which report solvability of only very small-size problems (Fu et al. 2005). In UC problem, the objective is often determined by $\sum_{t\in\mathscr{T}} f^p(\mathbf{p}_t) + \sum_{t\in\mathscr{T}, i\in\mathscr{I}} (S_i(\overline{\mathbf{u}}_{it}) + H_i(\overline{\mathbf{u}}_{it}))$, whereas the objective of ED problem only includes one of the followings: $f^p(\mathbf{p}_t)$, $f^v(\mathbf{V}_t)$, and $f^q(\mathbf{q}_t)$. It is common to include the DC power flow constraints (2) in the UC problem, ignoring the limits on the voltage magnitudes at buses, and the limits on phase shift

angles and tap positions of transformers. ED problem usually includes all the above-mentioned constraints for a single time period.

3 Unit Commitment

Unit commitment is the problem of finding an optimal up and down schedule and corresponding generation levels of a set of units over a planning horizon so that total cost of generation and transmission is minimized, the forecasted demand is satisfied, and a set of operating constraints such as upper and lower limits of generation, minimum up/down time limits, ramp up/down constraints, transmission constraints, and so forth, is observed. We give a mathematical formulation for the UC problem with linearized transmission constraints:

$$\min_{\mathbf{p},\mathbf{u},\boldsymbol{\theta}} \quad \sum_{t\in\mathscr{T},i\in\mathscr{I}} \left(f_i^p(p_{it}) + S_i(\overline{\mathbf{u}}_{it}) + H_i(\overline{\mathbf{u}}_{it})\right) \tag{13a}$$

$$\text{s.t.} \quad \widehat{PF}_b^t(\boldsymbol{\theta}_t) = \sum_{i\in\mathscr{I}_b} p_{it} - p_{bt}^d, \qquad b\in\mathscr{B},\ t\in\mathscr{T}, \tag{13b}$$

$$\widehat{FL}_{bb'}^t(\boldsymbol{\theta}_t) \le \overline{FL}_{bb'}, \qquad b,b'\in\mathscr{B},\ t\in\mathscr{T}, \tag{13c}$$

$$\underline{\theta}_{bb'} \le \theta_{bt} - \theta_{b't} \le \overline{\theta}_{bb'}, \qquad b,b'\in\mathscr{B},\ t\in\mathscr{T}, \tag{13d}$$

$$\underline{P}_i u_{it} \le p_{it} \le \overline{P}_i u_{it}, \qquad i\in\mathscr{I},\ t\in\mathscr{T}, \tag{13e}$$

$$p_{it} - p_{i,t-1} \le \overline{R}_i, \qquad i\in\mathscr{I},\ t\in\mathscr{T}, \tag{13f}$$

$$p_{i,t-1} - p_{it} \le \underline{R}_i, \qquad i\in\mathscr{I},\ t\in\mathscr{T}, \tag{13g}$$

$$u_{it} - u_{i,t-1} \le u_{i\tau}, \qquad \tau\in\{t,\ldots,\min\{t+\overline{M}_i-1,|\mathscr{T}|\}\}, \qquad i\in\mathscr{I},\ t\in\mathscr{T}, \tag{13h}$$

$$u_{i,t-1} - u_{it} \le 1 - u_{i\tau}, \qquad \tau\in\{t+1,\ldots,\min\{t+\underline{M}_i-1, \qquad |\mathscr{T}|\}\},\ \ i\in\mathscr{I},\ t\in\mathscr{T}, \tag{13i}$$

$$u_{it}\in\{0,1\}, \quad p_{it}\ge 0, \qquad i\in\mathscr{I},\ t\in\mathscr{T}, \tag{13j}$$

where $\widehat{FL}_{ij}^t(\boldsymbol{\theta}_t)$ is the simplified and linearized version of transmission flows (11).

UC problem is at the core of planning and operational decisions faced by ISOs, RTOs, and utility companies. Hence it has received a good deal of attention in the industry. In practice, UC is solved either in a centralized or decentralized manner (FERC 2006). In western and southern U.S. regions, most small utility companies generate and distribute their power in a decentralized manner, with California ISO (CAISO) in the west and Electric Reliability Council of Texas (ERCOT) in the south as two exceptions. In the Midwest, Pennsylvania New Jersey Maryland Interconnection (PJM) and Midwest Independent System Operator (MISO), and in Northeast, New York ISO (NYISO), and New England ISO (ISO-NE) handle most of the power system planning and transmission in a centralized manner. All these organizations

work in a centralized manner to solve their UC and ED problems (FERC 2006; Mukerji 2010; Ott 2010; Rothleder 2010).

Specifically, UC practices in PJM involve more than 600 utility companies in 13 states (Ott 2010). It manages 1210 generators with a total capacity of 164,905 MW, and it faces a peak load of 144,644 MW. PJM takes around 50,000 hourly demand bids from consumers one day ahead of the planning horizon. It then solves the day-ahead UC problem considering around 10,000 contingencies, and publishes generation schedules for the companies and locational marginal prices (LMPs) for the consumers. Consumers revise their bids based on the prices, and submit their final bids. Around 10,000 demand bids are submitted, and 8,700 of them are considered eligible. PJM updates the generation schedules and redoes its security analysis based on the final bids. In real-time, PJM solves a one-period security-constrained UC problem using full transmission model and turns on (off) some peaker units if the total planned amount that can be transmitted securely falls behind (exceeds) the demand.

The academic literature on the UC problem dates back to the 1960s (Garver 1962; Hara et al. 1966; Kerr et al. 1966), but developing exact solution approaches to real-size problems still remains as a challenge due to the combinatorial nature and high dimensionality of the problem. Researchers combine heuristic and exact algorithms along with decomposition techniques to cope with this challenge and obtain (near-)optimal solutions to the problem. Heuristic algorithms range from simple ranking schemes (Baldwin et al. 1959) to more sophisticated metaheuristic approaches (Mantawy et al. 1999). Exact algorithms include dynamic programming, mixed integer programming, Lagrangian relaxation, Benders decomposition, and, recently, column generation. The survey by Cohen and Sherkat (1987) presents several optimization algorithms applied to the problem. Viana (2004) summarizes metaheuristic algorithms; Padhy (2001) summarizes studies that use hybrid algorithms combining dynamic programming and various heuristic techniques; and the surveys (Padhy 2004; Salam 2007; Sheble and Fahd 1994) list a general mix of heuristic and exact algorithms. The book by Wood and Wollenberg (1996) addresses several operational and planning problems in the energy industry, including the UC problem.

3.1 Heuristic Algorithms

Heuristic methods for UC can be classified into two categories: pure heuristic approaches and complementary heuristic approaches. Pure heuristic approaches are stand-alone algorithms that aim to find feasible near-optimal solutions to the problem; whereas complementary heuristic approaches are developed to complement some other exact solution approaches that can prove bounds on the quality of the optimal solution but cannot find high-quality feasible solutions. We briefly summarize the pure heuristic approaches in this section, and leave the complementary

approaches to next sections where we present the exact algorithms that the heuristic algorithms complement.

Ranking heuristics that construct feasible solutions by starting up units based on priority lists are very common. Baldwin et al. (1959) uses such an approach forming the priority list based on average-full-load cost of units. Lee (1991) uses a similar approach forming the priority list based on cumulative utilization factors of units. The priority list heuristics can be translated into rules and executed as an expert system (Li et al. 1993; Mokhtari et al. 1987; Quyang and Shahidehpour 1990; Tong et al. 1991).

Mori and Matsuzaki (2001) propose a tabu-search heuristic algorithm based on a prioritized list of units. The local search starts from a priority list and is restricted to move to solutions that are close to the list. Other studies that employ tabu search include (Lin et al. 2002; Mantawy et al. 1998; Mantawy et al 2002). Maifeld and Sheble (1996) present a genetic algorithm approach with domain-specific mutation operations. Other studies using genetic algorithm include (Dasgupta and McGregor 1994; Kazarlis et al. 1996; Swarup and Yamashiro 2002; Yang et al. 1997). Simulated annealing (Mantawy et al. 1998; Zhuang and Galiana 1990), evolutionary programming (Chen and Wang 2002; Juste et al. 1999), artificial neural networks (Liang and Kang 2000; Sasaki et al. 1992), and ant colony search algorithms (Huang 2001; Sisworahardjo and El-Kaib 2002) are among the commonly used metaheuristic algorithm applied to the UC problem.

3.2 Dynamic Programming

Dynamic programming is one of the earliest methods applied to the UC problem. It decomposes a problem into small pieces (states), explores all combinations (stages) of these pieces, and recovers the overall optimum recursively. Different definitions of states and stages yield different dynamic programming algorithms for the UC problem. In one application (Ayoub and Patton 1971; Lowery 1966), the states are nested subsets of units, $\{1\}, \{1, 2\}, \ldots, \{1, 2, \ldots, |\mathscr{I}|\}$, and the stages are all possible total generation amounts of the units in such a subset. The cost of generating x MWs of power using $|\mathscr{I}|$ units, $f_{\mathscr{I}}(x)$, can be obtained by searching over all y's with $y < x$, i.e.,

$$f_{\mathscr{I}}(x) = \min_{y}\{f_{\{1,2,\ldots,|\mathscr{I}|-1\}}(x - y) + g_{\mathscr{I}}(y)\},$$

where $g_{\mathscr{I}}(y)$ is the cost of generating y MWs of power on unit $|\mathscr{I}|$. In another implementation (Cohen and Sherkat 1987), the state space consists of up and down combinations of all units over all time periods, leading to $2^{|\mathscr{I}|} \times |\mathscr{T}|$ states.

The dynamic programming approach developed by Snyder (1987) is one of the earliest successful dynamic programming implementation. The algorithm features a classification and prioritization of units to reduce the number of states. The authors address the problem at San Diego Gas and Electric System with 30 generators. A

similar approach is proposed by Hobbs et al. (1988), where the authors use selection techniques for choosing the most promising states and implement approximate ED subroutines.

Different priority lists and different strategies to select a set of units from a priority list have been adopted with dynamic programming to limit the search space, which usually comes at a price of loosing the optimality. These strategies include dynamic programming with sequential combination (DP-SC), dynamic programming with truncated combination (DP-TC), and variable window-truncated dynamic programming (DP-VW). In DP-SC strategy, the least costly units are committed first, and the most costly units are committed last. The search space is reduced by considering the combinations of units sequentially (Pang et al. 1981). In DP-TC strategy, a small portion of the solution space is considered within the priority list ordering (Pang and Chen 1976). In DP-VW strategy, a window and a threshold are defined to determine which units are automatically committed, which units are considered for commitment, and which are not considered at all (Quyang and Shahidehpour 1992). Pang et al. (1981) compare the performances of four UC methods, three of which are based on dynamic programming including the strategies DP-SC and DP-TC.

Li et al. (1997) introduce a new dynamic programming approach based on a decommitment procedure. From an initial schedule of all online units committed over the planning horizon, decommitment of units is executed sequentially using the dynamic programming principle until no more reduction in the cost or no more change in the schedule is achieved over two consecutive iterations. Hobbs et al. (1988) develop a realistic UC solution approach based on dynamic programming. Siu et al. (2001) develop a real-size hydrodynamic UC and loading model for the British Columbia Hydro Power Authority.

3.3 Lagrangian Relaxation

The main drawback of dynamic programming is its high dimensionality and its inability to handle constraints that affect multiple periods, such as ramping constraints. Number of states in dynamic programming increases exponentially by the problem size. Due to these drawbacks, Lagrangian relaxation gained increased popularity in 1980s. The main idea in Lagrangian relaxation is to dualize (relax) the complicating and linking constraints by penalizing and appending their violations to the objective function, and to employ an iterative procedure that finds the best penalties for the dualized constraints so that the new relaxed problem (Lagrangian problem) gives the tightest bound for the original problem. The problem of searching for the best penalties is called the Lagrangian dual problem and uses one of several alternative algorithms, including the well-known subgradient algorithm. The Lagrangian problem, obtained by dualizing the complicating constraints, usually decomposes into simple small subproblems. In UC setting, it is common to relax the demand constraints (or the transmission constraints), the reserve constraints, and the fuel constraints, and to decompose the original UC problem into independent single-unit UC subproblems

that are usually solved by dynamic programming. Ramping constraints, if exist, hinder the ability to apply dynamic programming to these subproblems; therefore, it is a common practice to relax the ramping constraints as well.

Lagrangian relaxation, however, comes with two limitations: the duality gap and the inability to maintain the primal feasibility. In non-convex optimization problems, such as the UC problem, the optimal solution value of the Lagrangian dual is not equal to the optimal solution value of the original problem. It is necessary to take actions such as branching to close this duality gap. For large UC problems, it is observed that the duality gap is less than 0.5 % and goes to zero as the problem size gets larger (Lauer et al. 1982). Another drawback of Lagrangian relaxation is its inability to produce primal feasible solutions automatically. The optimal primal solutions to the Lagrangian dual is usually infeasible to the original problem, and thus it is necessary to resort to some repair heuristics techniques.

There is an ample literature employing the above idea to the UC problem. The study by Muckstadt and Koenig (1977) is one of the first studies to do so. The authors use Lagrangian relaxation in a branch-and-bound framework instead of the common linear programming relaxation, and obtain significant improvement of computational efficiency, due to smaller duality gaps, as compared to the linear programming relaxation. They relax the demand and reserve constraints and solve the resulting subproblems by dynamic programming.

Cohen and Wan (1987) apply Lagrangian relaxation to a UC problem with fuel constraints. They relax the fuel constrains, demand constraints, and reserve constraints, and update the penalties for each constraint sequentially to reduce the computation time. Aoki et al. (1987) employ a similar approach but update the penalties simultaneously. Tong and Shahidehpour (1990) consider a UC problem with thermal and hydro units. The authors append an additional constraint to the problem so that at each iteration of the algorithm the solutions to subproblems do not change significantly. At an iteration of the algorithm, if the solution is feasible, a post-processor based on linear programming is applied to assure that the dispatch with the minimum cost is attained.

Lauer et al. (1982) update the Lagrangian multipliers using the second derivative information as opposed to the commonly used subgradient algorithm. Bertsekas et al. (1983) resolve the Lagrangian dual problem repeatedly until an optimality gap within a tolerance is obtained. Merlin and Sandrin (1983) solve a practical UC problem arising at Electricite de France using Lagrangian relaxation.

To recover a primal feasible solution from the Lagrangian dual solution, Zhuang and Galiana (1988) suggest increasing the penalties associated with the most violated constraint and repeating the process until a feasible solution is obtained. Bard (1988) dualizes the demand and reserve constraints and handles the resulting subproblems by dynamic programming with discretized generation variables. The study successfully solves problems with 100 units and 48 time periods within one % optimality gap. Guan et al. (1996) dualize ramping constraints in addition to the demand constraint to be able to solve the subproblems by dynamic programming. Lai and Baldick (1999) use Lagrangian relaxation to deal with ramping constraints that differentiate the startup and shutdown periods from the other periods.

Baldick (1995) proposes a constructive and a post-processing method to produce and refine primal feasible solutions in the case of a generalized UC problem that includes power flow constraints, line flow limits, voltage limits, and total fuel and energy limits. Birge and Takriti (2000) develop a Lagrangian relaxation algorithm based on relaxing the demand and reserve constraints and propose a procedure based on an integer program that takes subproblem solutions from all previous iterations of the algorithm and produces primal feasible solutions to the problem. Other studies using Lagrangian relaxation are (Ruzic and Rajakovic 1991; Virmani et al. 1989).

3.4 Mixed Integer Programmming and Branch-and-Bound

Garver (1962) is one of the earliest studies that apply cut generation to the linear programming relaxation of the UC formulation, which is the basic idea in today's highly sophisticated mixed integer programming solvers. The author generates constraints that are valid to the integer variables but are violated by the linear programming relaxation solution. An integer programming approach proposed by (Dillon et al. 1978) addresses the UC problem of hydro-thermal systems with reserve requirements. It is one of the earliest papers that can solve real-life problems with 20 units. The authors develop two sets of valid inequalities that are globally valid to the problem and use these inequalities in a branch-and-bound algorithm. Cohen and Yoshimura (1983) present a novel branch-and-bound approach assuming that units can be turned on and off only once throughout the planning horizon. Bond and Fox (1986) introduce an algorithm based on integration of mixed integer linear programming and dynamic programming. A mixed integer linear programming problem determines a feasible set of units at each scheduling point, while a dynamic programming problem identifies promising scheduling routes.

Lee et al. (2004) develop facet defining inequalities for a projection of minimum up and down time constraints in the UC problem without startup costs. Rajan and Takriti (2005) develop a similar set of inequalities for the problem with the startup costs, and show that their inequalities are stronger than those in Garver (1962) and stronger than those in Lee et al. (2004) in the case of a UC problem with the startup costs. Chang et al. (2004) and Li and Shahidehpour (2005) develop a similar set of inequalities for the UC problem with the shutdown variables. Hedman et al. (2009) compare and contrast the inequalities developed in Chang et al. (2004); Garver (1962); Lee et al. (2004); Li and Shahidehpour (2005); Rajan and Takriti (2005).

3.5 Benders Decomposition

Benders decomposition method separates the solution of UC and ED problems. A master problem (UC problem) is solved to obtain unit schedules, and these schedules are supplied to a set of subproblems (ED problems). Subproblems are solved to

check the feasibility and optimality of the unit schedules. If either feasibility or optimality is violated, constraints that eliminate the current schedule is returned to the master problem. The master problem and the subproblems are solved iteratively until optimal unit schedules that satisfy the ED constraints are obtained. Any ED solver can be applied to solve the subproblems. The main drawback of Benders decomposition approach is to solve the master problem, which is still a large-scale integer optimization problem.

Muckstadt and Wilson (1968) solve a stochastic UC problem with transmission constraints using a Benders decomposition approach. It is one of the earliest studies to address the stochasticities in the problem. Turgeon (1978) applies Benders decomposition and addresses the solution of only the master problem. The author solves the master problem by a Lagrangian relaxation approach implemented within a branch-and-bound framework. Baptisella and Geromel (1980) is one of the earliest studies that use stochastic programming modeling approach in UC problem. The authors address a problem with both hydro and thermal units and with stochastic load parameters. The objective is to minimize the operating cost of thermal units and the cost of expected unsatisfied load. The model also includes a spinning reserve constraint to cover unexpected failure of the largest generator. The authors use a Benders decomposition approach where the master problem optimizes the thermal units and meets the demand, subproblems solve for hydro schedules. The master problem is further decomposed by Lagrangian relaxation to obtain fast solutions. Habibollahzadeh and Habibollahzadeh and Bukenko (1986) apply Benders decomposition such that the master problem contains only integer variables and subproblems include the constraints for hydro and thermal units. They do not use the minimum up and down time constraints, but include a constraint that allows only one commitment per day for each unit. Ma and Shahidehpour (1998) deal with UC with transmission constraints and phase-shifter transformers. After applying Benders decomposition, the master problem becomes a pure UC problem and subproblems become ED problems with transmission constraints. The authors solve the master problem by augmented Lagrangian method. Fu et al. (2005) apply Benders decomposition to security-constrained UC problem, and solve the master problem with augmented Lagrangian and dynamic programming. Recently, Wu and Shahidehpour (2010) improve Benders cuts to solve the UC problem with transmission constraints.

4 Unit Commitment Under Uncertainty

What makes electricity unique and different from other commodities is that it is not storable and its demand has to be met in real time. This constitutes a challenging problem to practitioners and academicians, as the uncertainty in the demand, in unit and network availabilities, and in renewable generation makes it impossible to forecast the future status of the system. Although after deregulation of the power market, demand uncertainty started to disappear due to power purchase agreements and power contracts, the advent of renewable generation with its intermittent nature

creates uncertainties in the net demand. The traditional way of handling these uncertainties is to impose an operating reserve capacity requirement into the model with the hope of reacting to such uncertain events. By the advances in computational and algorithmic tools, recently stochastic programming and robust optimization techniques have started to appear to address these uncertainties.

4.1 Operating Reserves

Although there is no consensus on the very exact definition of operating reserves, the basic functionality of such a reserve capacity requirement is to assure that in the case of unexpected demand increases or unit or transmission outages, the system remains in a stable condition. Two types of operating reserves are studied: Spinning reserve is the unused capacity of an online unit that can respond to an uncertain event immediately; non-spinning reserve is unused capacities of (possibly offline) units that can respond to emergency within half an hour of its occurrence. Various sets of constraints have been used to impose such requirements. A typical set of constraints for spinning reserve reads:

$$\sum_{i \in \mathscr{I}} r_{it} > R_t^S, \qquad t \in \mathscr{T}, \tag{14}$$

$$r_{it} + p_{it} - p_{i,t-1} \leq \overline{R}_i, \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{15}$$

$$r_{i,t-1} + p_{i,t-1} - p_{it} \leq \underline{R}_i, \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{16}$$

$$r_{it} + p_{it} \leq \overline{P}_i u_{it}, \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{17}$$

$$r_{it} > 0, \qquad i \in \mathscr{I},\ t \in \mathscr{T}, \tag{18}$$

where r_{it} denotes the spinning reserve capacity of unit i at time period t, and R_t^S denotes the spinning reserve requirement at time period t. Note that the ramping constraints and active power limit constraints are modified accordingly. A more simplistic formulation is sometimes preferred:

$$\sum_{i \in \mathscr{I}} \overline{P}_i u_{it} - p_{it} > R_t^S, \qquad t \in \mathscr{T}. \tag{19}$$

On the other hand, non-spinning reserve requirement does not interfere with the ramping constraints or up and down statuses of units:

$$\sum_{i \in \mathscr{I}_P} \overline{P}_i - p_{it} > R_t^N, \qquad t \in \mathscr{T}, \tag{20}$$

where $\mathscr{I}_P$ is the subset of the set of units, called the peaker units that can be ramped up in half an hour, and R_t^N is the non-spinning reserve requirement at time period t.

Zhang et al. (2010) and Yong et al. (2009) give a similar representative formulation for the operating reserve requirement.

4.2 Stochastic Unit Commitment

Stochastic programming is a widely used tool to model uncertainties in the UC problem. Two widely modeled uncertainties are the load uncertainty (Takriti and Birge 1996) and unit outages (Carpentier et al. 1996). Stochastic optimization uses a probability distribution for the uncertain variables in the optimization problem. The objective function in a stochastic optimization model is usually the first two moments, expectation or variance, of the cost function. The constraints are to be satisfied with a probability. If the uncertain variable is a stochastic process that evolves over time, it is common to use a dynamic model that has the flexibility of adjusting the decision variables based on how the uncertain variable realizes over time. Such stochastic models are called multi-stage models, as opposed to two-stage models where decisions are made before realizing the uncertainty, and these decisions do not change throughout the planning horizon no matter how the uncertainty evolves. If the decision maker is allowed to take a recourse action when the constraints of the model is not satisfied after realizing the uncertainty, the model is called a stochastic recourse model; otherwise, it is called a chance-constrained model. We give a multi-stage stochastic chance-constrained model for the UC problem with uncertain loads, or the renewable power, having a discrete probability distribution.

We assume that the uncertainty evolves as a discrete time stochastic process with a finite probability space. We represent the information structure as a rooted scenario tree where the nodes n ($n \in \mathcal{N}$) in level t ($t \in \mathcal{T}$) of the scenario tree constitute the states of the world that can be distinguished by the information available up to time period t (Ahmed et al. 2003; Singh et al. 2009). The set of *leaf nodes*, $\mathcal{N}_L$ ($\mathcal{N}_L \subset \mathcal{N}$), contains the nodes without any successor. The root node is the node without any predecessor. In general, $n(\tau) \in \mathcal{N}$ represents the τ th predecessor of node n. The level of the root node is zero, and in general the level of a node, t_n, is defined such that $n(t_n)$ is the root node. The tree has a depth of $|\mathcal{T}|$, and all leaf nodes have a level of $|\mathcal{T}|$. By convention, $n(0)$ is the node n itself, and $n(\tau),\ \tau > t_n$, is an empty set.

The root node has an occurrence probability of one. For each node $n \in \mathcal{N}$, p_n denotes the probability that the corresponding state of the world occurs given that its predecessor, $n(1)$, has occurred; and π_n denotes the unconditional probability that the corresponding state occurs, i.e., $\pi_n = \prod_{\tau=0}^{t_n-1} p_{n(\tau)}$. There is a one-to-one matching between the leaf nodes of the scenario tree and the scenarios. Given a leaf node, $n \in \mathcal{N}_L$, a $|\mathcal{T}|$-tuple $[n(|\mathcal{T}|), n(|\mathcal{T}|-1), \ldots, n(1), n]$ represents a scenario with probability of occurrence equal to π_n. Two scenarios sharing the same state of the world at time periods $1, \ldots, \tau$, for some $\tau < |\mathcal{T}|$, have to observe the same set of decision variables in the optimization model, in order to make sure that the model does not cheat by foreseeing (anticipating) the future. Using only a single set

of decision variables for each node guarantees such a non-anticipativity property, and yet keeps the model size small as compared to using a separate set of decision variables for each scenario and for each time period and setting the variables equal to each other (Lulli and Sen 2004; Takriti et al. 2000).

The stochastic UC model (21) is an extension of model (13), where nodes n in the scenario tree replaces the time periods t in the deterministic model (13), and $\overline{\mathbf{u}}_{in} = (u_{i,t_n}, u_{i,t_n+1}, \ldots, u_{in})$.

$$\min_{\mathbf{p},\mathbf{u},\boldsymbol{\theta}} \sum_{n\in\mathcal{N}, i\in\mathcal{I}} \pi_n \left(f_i^p(p_{in}) + S_i(\overline{\mathbf{u}}_{in}) + H_i(\overline{\mathbf{u}}_{in}) \right) \tag{21a}$$

$$\text{s.t. } \widehat{PF}_b^n(\boldsymbol{\theta}_n) = \sum_{i\in\mathcal{I}_b} p_{in} - p_{bn}^d, \qquad b \in \mathcal{B},\ n \in \mathcal{N}, \tag{21b}$$

$$\widehat{FL}_{bb'}^n(\boldsymbol{\theta}_n) \le \overline{FL}_{bb'}, \qquad b, b' \in \mathcal{B},\ n \in \mathcal{N}, \tag{21c}$$

$$\underline{\theta}_{bb'} \le \theta_{bn} - \theta_{b'n} \le \overline{\theta}_{bb'}, \qquad b, b' \in \mathcal{B},\ n \in \mathcal{N}, \tag{21d}$$

$$\underline{P}_i u_{in} \le p_{in} \le \overline{P}_i u_{in}, \qquad i \in \mathcal{I},\ n \in \mathcal{N}, \tag{21e}$$

$$p_{in} - p_{i,n(1)} \le \overline{R}_i, \qquad i \in \mathcal{I},\ n \in \mathcal{N}, \tag{21f}$$

$$p_{i,n(1)} - p_{in} \le \underline{R}_i, \qquad i \in \mathcal{I},\ n \in \mathcal{N}, \tag{21g}$$

$$u_{i,n(\tau)} - u_{i,n(\tau+1)} \le u_{in}, \qquad \tau \in \{0, 1, \ldots, \min\{\overline{M}_i - 1, t_n\}\},\ i \in \mathcal{I},\ n \in \mathcal{N}, \tag{21h}$$

$$u_{i,n(\tau+1)} - u_{i,n(\tau)} \le 1 - u_{in}, \qquad \tau \in \{0, 1, \ldots, \min\{\underline{M}_i - 1, t_n\}\},\ i \in \mathcal{I},\ n \in \mathcal{N}, \tag{21i}$$

$$u_{in} \in \{0, 1\}, \quad p_{in} \ge 0, \qquad i \in \mathcal{I},\ n \in \mathcal{N}. \tag{21j}$$

Model (21) is a commonly used model in stochastic UC and is adopted by(Gröve-Kuska and Römisch 2005; Nowak and Romisch 2000; Takriti and Birge 1996; Koc and Kalagnanam 2012) to model the uncertainty in the demand. Koc and Kalagnanam (2012) solve model (21) using a parallel branch-cut-price framework that combines column and cut generation algorithms, primal heuristics and two parallelization approaches. Takriti et al. (2000) use a similar model with a recourse, i.e., the model has a flexibility of meeting the demand both with power generation and power purchase, and minimize expected power generation and purchase costs. They solve the problem using a Lagrangian relaxation method. Zhang et al. (2010) use a similar approach that also models transmission constraints. Their objective function includes generation cost, reserve shortage cost, and load shedding cost. They develop a heuristic algorithm to solve the resulting large-scale problem. Koc and Ghosh (2012) approach from a different perspective. They develop a scenario reduction algorithm that reduces the size of model (21) so that the reduced model can be solved faster by any solution technique, and yet the reduced model is close to the original model as much as possible, in terms of optimal solutions, and the optimal value.

The stochastic model that models unit availabilities (or outages) is very similar to model (21), except that in this case the scenarios specify the availability of units instead of the demand realizations. Carpentier et al. (1996) solve such a model with augmented Lagrangian method. Their objective function is the expected operating cost and the expected penalty for unmet demand. Wu et al. (2007) consider a

multi-stage stochastic security-constrained UC problem that incorporates both the uncertainty in the demand and in unit availabilities.

Two-stage stochastic optimization models are also common in UC setting. Liu et al. (2010) develop an extended Benders decomposition algorithm for a two-stage stochastic UC problem. Xiong and Jirutitijaroen (2011) consider a two-stage UC problem and use a multi-cut version of Benders decomposition as a solution approach.

4.3 Robust Unit Commitment

Robust optimization deals with uncertain parameters in an optimization problem assuming no knowledge on the parameters except an uncertainty set in which the parameters reside. It optimizes against all possible values of the uncertain parameters, and gives a solution that optimizes the worst performance of the model under all possible realizations of the uncertain parameters. This approach models UC problems as a two-stage formulation, as compared to the multi-stage stochastic UC formulation:

$$\min_{\mathbf{u}} \sum_{t\in\mathscr{T},i\in\mathscr{I}} \left(S_i(\overline{\mathbf{u}}_{it}) + H_i(\overline{\mathbf{u}}_{it}) + \max_{p^d_{bt}\in\mathscr{D}} \min_{\mathbf{p},\boldsymbol{\theta}} \sum_{t\in\mathscr{T},i\in\mathscr{I}} f_i^p(p_{it}) \right) \tag{22a}$$

$$\text{s.t.} \quad (13b) - (13j), \tag{22b}$$

where $\mathscr{D}$ denotes the uncertainty set of active power demand. Different assumptions on this uncertainty set results in different models. Bertsimas et al. (2013) consider an uncertainty set with $\mathscr{D} = \mathscr{D}^1 \times \cdots \mathscr{D}^t \times \cdots \mathscr{D}^{|\mathscr{T}|}$, where $\mathscr{D}^t = \{p^d_{bt} : \sum_{b\in\mathscr{B}} \frac{|p^d_{bt} - \bar{p}^d_{bt}|}{\hat{p}^d_{bt}} \le \nabla_t, \quad p^d_{bt} \in [\bar{p}^d_{bt} - \hat{p}^d_{bt}, \bar{p}^d_{bt} + \hat{p}^d_{bt}], b \in \mathscr{B}\}$. The demand value p^d_{pt} is allowed to take any value within the symmetric set $[\bar{p}^d_{bt} - \hat{p}^d_{bt}, \bar{p}^d_{bt} + \hat{p}^d_{bt}]$, but a limit of ∇_t is assumed on the sum of normalized demands over all busses. The authors consider a UC problem with transmission constraint. Zhang and Guan (2011) develop a similar approach to UC problem without transmission and ramping constraints. In their model, the uncertain demand is linked over time periods, i.e., $\mathscr{D} = \{p^d_{bt} : \sum_{t\in\mathscr{T},b\in\mathscr{B}} \nabla_t p^d_{bt} \le \nabla_0, \quad p^d_{bt} \in [\overline{p}^d_{bt}, \underline{p}^d_{bt}], \quad b \in \mathscr{B}, \ t \in \mathscr{T}\}$.

Street et al. (2011) develop a robust optimization approach to N-k security-constrained UC problem, where any k of the n units can fail simultaneously and the system still guarantees to satisfy the demand with the remaining units. The authors develop a deterministic equivalent of the problem that can be solved by linear mixed integer programming solvers.

5 Economic Dispatch

Economic dispatch is the problem of determining the most efficient, low-cost, and reliable operation of a power system by dispatching the available electricity generation resources to the load on the system via transmission lines. The primary objective of economic dispatch is to minimize the total cost of generation while satisfying the physical constraints and operational limits. This problem is solved at a smaller time scale than the UC problem—typically five to fifteen minutes. The ED problem plays an important role in power system analysis (Glover et al. 2008; Wood and Wollenberg 1996), especially in planning, operation, and control of power systems.

The ED problem is formulated as a nonlinear programming problem, whose conventional constraints include either the DC or the AC power flow equations, physical limits of the control variables, physical limits of the state variables, and other limits such as power factor limits (Glavitsch and Bacher 1991). All electricity is transferred to the points of demand using AC transmission lines, except for a small percentage which is transported in DC form using high-voltage DC links (Acha et al. 2005). In this paper, we focus on AC power flow constraints for the ED problem. We define $\mathscr{G}$ as the set of active generators obtained from the UC problem. Considering only the active generators, the single-stage ED problem becomes

$$\min_{\mathbf{p},\mathbf{q},\mathbf{V},\boldsymbol{\theta},\mathbf{t},\boldsymbol{\phi}} \quad \sum_{i\in\mathscr{G}} f_i^p(p_i) \tag{23a}$$

$$\text{s.t.} \quad PF_b(\mathbf{V}) = \sum_{i\in\mathscr{I}_b} p_i - p_b^d, \qquad b \in \mathscr{B}, \tag{23b}$$

$$QF_b(\mathbf{V}) = \sum_{i\in\mathscr{I}_b} q_i - q_b^d, \qquad b \in \mathscr{B}, \tag{23c}$$

$$FL_{bb'}(\mathbf{V}) \le \overline{FL}_{bb'}, \qquad b, b' \in \mathscr{B}, \tag{23d}$$

$$\underline{P}_i \le p_i \le \overline{P}_i, \qquad i \in \mathscr{G}, \tag{23e}$$

$$\underline{Q}_i \le q_i \le \overline{Q}_i, \qquad i \in \mathscr{G}, \tag{23f}$$

$$\underline{V}_b \le |V_b| \le \overline{V}_b, \qquad b \in \mathscr{B}, \tag{23g}$$

$$\underline{\theta}_{bb'} \le \theta_b - \theta_{b'} \le \overline{\theta}_{bb'}, \qquad b, b' \in \mathscr{B}, \tag{23h}$$

$$\underline{t}_l \le t_l \le \overline{t}_l, \qquad l \in \mathscr{L}, \tag{23i}$$

$$\underline{\phi}_l \le \phi_l \le \overline{\phi}_l, \qquad l \in \mathscr{L}. \tag{23j}$$

Note that we omit the time period index t from the formulation. In today's market, only the cost of active power is concerned. Quadratic costs are widely used, but in the bidding electricity market system, piecewise linear costs are preferred. Other forms of the objective function are considered as well, such as the reactive power loss on transmission lines and the total system active power losses. The difficulty of the problem mainly comes from the highly nonlinear and non-convex constraints (23b), (23c), and (23d). The ED problem becomes a mixed integer nonlinear pro-

gramming problem when discrete control variables such as transformer taps, shunt capacitor banks, and other flexible AC transmission system (FACTS) devices are taken into account (Acha et al. 2005). Furthermore, if transient stability constraints are considered as the dynamic stability conditions, the problem consists of a set of large-scale differential-algebraic equations (Gan et al. 2000; Jiang and Geng 2010). In this section, we focus on the purely continuous variable single-stage ED problem, relaxing the integrality of discrete control variables such as transformer taps and shunt capacitor banks.

In the early 1960s, Carpentier originally introduced the ED problem and made the first attempt to solve it (Carpentier 1962). Since then, there has been a lot of studies to address the solution of the problem including local centralized approaches, distributed algorithms, and global optimization techniques.

5.1 Local Centralized Approaches

Because of the non-convexity of the problem and the need for fast and robust algorithms, most solution approaches attempt to find a local optimum to the ED problem that satisfies the first-order optimality conditions. Popular numerical centralized techniques that do not decompose the problem include successive linear/quadratic programming, trust-region-based methods, Lagrangian-Newton methods, and interior-point methods.

Successive linear/quadratic programming. This approach approximates the objective function by a linear or quadratic function and successively linearizes the constraints. Wells (1968) utilizes this idea to derive a sequence of linear programming to solve the ED problem with security constraints. Contaxis et al. (1986) decompose the ED problem into real and reactive subproblems and solve these two subproblems using quadratic programming in each iteration. The algorithm by Amerongen (1988) rigorously linearizes the necessary Karush-Kuhn-Tucker conditions, and then transforms them into a sequence of related quadratic programming problems. The algorithm exploits the sparsity structure of the problem to speed up the computations by using explicit reduction of some of the variables. One intrinsic disadvantage of these methods is their poor computational results and convergence rates. They often fail to handle large-scale problems.

Trust-region based methods. Trust-region methods belong to a class of optimization algorithms that minimize an (typically quadratic) approximation of the objective function within a closed region, called the trust-region. Various methods differ in the way to choose the trust-region. Min and Shengsong (2005) propose a trust-region interior-point algorithm with two iterations: a main iteration and a linear programming (LP) inner iteration. The algorithm linearizes the ED problem to form a trust-region subproblem in the main iteration. The LP inner iteration solves the trust-region LP subproblem by the multiple centrality corrections primal-dual interior-point method. The trust-region controls the linear step size. The authors show that their algorithm is superior to successive linear programming methods. The

method of Sousa and Torres (2007, 2011) is based on an infinity-norm trust-region approach, and uses interior-point methods to solve the trust-region subproblems. The convergence robustness is tested and verified by using different starting points.

Lagrangian-Newton methods. Sun et al. (1984) propose one of earliest Lagrangian-Newton method approaches that solve the Lagrangian function by minimizing a quadratic approximation. Santos and Costa (1995) study an augmented Lagrangian method that incorporates all the equality and inequality constraints into the objective function. The first-order optimality conditions are achieved by Newton's method together with a multiplier update scheme. Baptista et al. (2005) consider the application of logarithmic barrier method to voltage magnitude and tap-changing transformer variables and treat the other constraints by an augmented Lagrangian method. Wang et al. (2007) handle inequality constraints in the context of the quadratic penalty method by using squared additional variables to form a sequence of unconstrained optimization problems. The authors use a trust-region method based on a 2-norm subproblem to solve the unconstrained problems.

Interior-point methods. A large number of efficient methods based on the Karush-Kuhn-Tucker necessary conditions use interior-point methods (Capitanescu et al. 2007; Chiang et al. 2009; Jabr 2003; Torres and Quintana 1998; Wang et al. 2007). Interior-point methods (IPMs) have been widely applied to solve the ED problem in the last decade, owing to its well-known excellent properties for large-scale optimization problems. In particular, its local quadratic convergence is established, and a class of polynomial-time interior-point algorithms have been designed. The IMPs transform the inequality constraints into equality constraints by introducing nonnegative slack variables, and then, typically, treat the slack variables via a logarithmic barrier term. The main idea in these algorithms is to exploit two relatively standard powerful techniques: homotopy logarithmic barrier function to deal with inequality constraints and Newton's method for a system of equations. The work by Jiang et al. (2010) additionally proposes an implementation of the automatic differentiation technique for the ED problem.

Among these local approaches, IPMs often show the best performance; and, they are, and will likely be in the future, one of the most practical approaches designed for the very large-scale, real-world electric networks. The-state-of-the-art implementations of general-purpose interior-point methods such as KNITRO (Byrd et al. 2006) and IPOPT (Wächter and Biegler 2006) can handle networks containing tens of thousands of nodes, although they sometimes do not converge to a feasible solution for some different initial points.

5.2 Distributed Algorithms

When the size of electric networks grows tremendously, solving the ED problem in a centralized manner might not be practical. Some approaches are devoted to decompose the problem into smaller subproblems, each of which can be solved independently and effectively by different entities in the network. Motivated by the

distributed multi-processor environments, Kim and Baldick (1997, 2000) present some decomposition algorithms based on the auxiliary problem principle, the proximal point method, and the alternating direction method of multipliers. They split the power grid into a number of separate regions. By duplicating the variables in overlapping regions, they are able to solve the ED problem by these methods in a distributed way. Lam et al. (2011) show that for problems with special structures such as trees in distribution networks, the semi-definite programming (SDP) relaxation has a zero duality gap, and therefore they propose primal and dual decomposition algorithms to solve the dual problem.

5.3 Global Optimization

Economic dispatch is an NP-hard non-convex, nonlinear optimization problem (Lavaei and Low 2012), and it is in general difficult to be solved to optimality. Apart from the above-mentioned local methods, some researchers have attempted to find a convex formulation for the problem. In Jabr (2006), the author shows that the load flow problem of a radial distribution system can be modeled as a convex optimization problem in the form of conic programming. In a meshed network, nevertheless, the convexity cannot be derived, and the problem is formulated in an extended conic quadratic format (Jabr 2008). To the best of our knowledge, Bai et al. (2008) are among the first to consider solving the ED problem by semidefinite programming, though they do not make a theoretical justification on the solution quality. Recently, Lavaei (2011); Lavaei and Low (2012); Sojoudi and Lavaei (2012) have proposed other semidefinite programming relaxations for the ED problem. They introduce a sufficient condition based on the solution of the dual for a zero duality gap that is satisfied by a range of power grid test instances including all IEEE benchmark systems, and expectedly every practical power system; hence it guarantees that semidefinite programming can solve the problem to optimality. Obviously, it would be desirable to prove that the duality gap is necessarily zero, but, to note, nobody has been able to do this. However, it is known that one of drawbacks of semidefinite programming is the lack of its scalability to solve very large-scale problems; the current interior-point methods for SDP can only handle problems with size up to several hundreds (Wolkowicz et al. 2000). For this approach to be practical, more research on its solution methodology is needed. Phan (2012) investigates a Lagrangian dual problem based on the 2-norm trust-region subproblem for solving the ED problem in rectangular form. When the strong duality does not hold for the dual, the author proposes two classes of branch-and-bound algorithms that guarantee to solve the problem to optimality. The lower bound for the objective function is obtained by the Lagrangian duality, and the feasible set subdivision is based on the rectangular or ellipsoidal bisection. However, the author remarks that no duality gap is observed for any test problems.

5.4 Economic Dispatch with Renewable Resources

With the advent of the Smart Grid, the infrastructure for energy supply generation and transmission is experiencing a transition from the current centralized system to a decentralized one. The responsiveness and flexibility envisioned for the Smart Grid provide additional advantages in facing the significant new challenges of integrating distributed and intermittent generation capability, such as small generators and renewable energy sources (wind, solar, etc.), at a scale that current grid technology is finding hard to achieve (Cheung et al. 2010; Li et al. 2007). This is becoming more critical as renewable energy technologies are playing an increasingly visible role in the portfolio mix of electricity generation.

Renewable generations such as wind and solar have negligible operational costs, in the hourly time scale, and thus should be the first generator to be dispatched. Indeed, regulations in multiple US states require the use of wind power if it is being generated. However, the intermittent nature of output from wind turbines due to weather conditions is often seen as a potential obstacle to dispatching wind power in the classical sense. Hence, we often model the wind power as a non-dispatchable, variable generation source that is connected in an always-on state to the grid.

Forecasting near-term wind availability and velocity is an imperfect science with significant variability between the forecasted and the realized generation. In Dragoon and Milligan (2003), the authors consider integrating wind power production into existing dispatch models, and analyze the effect of the forecast errors of wind power production on the incremental reserve requirements and imbalance costs. The agency charged with controlling the smooth operation of the grid will require that the uncertainty associated with utilizing non-renewable sources be hedged against. This problem is often addressed by balancing energy provided by non-dispatchable sources, such as wind and photovoltaic units, with quickly dispatchable, albeit costly, sources, such as small hydro and micro turbine units. This problem has been studied in various levels of sophistication starting from individual end users up to local utilities. In particular, a balancing approach to achieve overall dispatchability in a distributed generation network is presented in Xue et al. (2007), which consequently converts a group of small distributed generations into a large logical generation station.

Another stream of research incorporates the uncertainty in setting or adjusting the dispatch and transmission decisions, which has the effect of dispatching additional capacity to hedge against the risk of a large unforeseen shortfall in total supply. The stochastic ED can be solved by imposing a set of risk constraints, in the form of chance constraints in Fu and McCalley (2001) or mean-excess constraints in Ghosh et al. (2011), to balance risk of shortfalls due to uncertain generation against cost of provisioning corrective generation sources such as peakers. The study in Brini et al. (2009) considers an economic environmental dispatching model where wind and solar energy are both included but constrained to be no more than 30% of the total dispatch capacity. Hatami et al. (2009) propose a stochastic programming framework to determine the optimal procurement of interruptible load in order to minimize the risk of a shortfall over multiple periods.

Phan and Ghosh (2011) propose a two-stage stochastic formulation to address the wind-generation uncertainty. The two stages model dispatching and transmission decisions that are made on subsequent time periods separated by a small time period, such as fifteen minutes or an hour. Certain generation decisions are made only in the first stage; and the second stage realizes the actual renewable generation, where the uncertainty in renewable output is captured by a finite number of scenarios. They present two outer approximation algorithms, and show that under certain conditions the sequence of optimal solutions obtained under both alternatives has a limit point that is a globally-optimal solution to the original two-stage non-convex problem.

6 Security-Constrained Economic Dispatch

We investigate some mathematical formulations for the security-constrained economic dispatch (SCED) problem and deterministic solution methods in the literature. Unlike the classical ED problems, the SCED problems take into account both the pre-contingency (base-case) constraints and post-contingency constraints. Our review focuses on the widely used N-1 contingency; that is, even if a single component, such as a transmission line, generator, or transformer, is out of service, the power system should still satisfy the load requirements without any operating violations.

The first type of SCED formulation is the preventive SCED (Alsac and Stott 1974), where some objective function is minimized by acting only on the base-case (contingency-free) control variables subject to both normal and abnormal operating constraints with one of the contingencies. The problem is modeled as

$$\begin{aligned} \min_{\mathbf{x}_0,\ldots,\mathbf{x}_c,\mathbf{u}_0} \quad & f(\mathbf{x}_0,\mathbf{u}_0) \\ \text{s.t.} \quad & \mathbf{g}_k(\mathbf{x}_k,\mathbf{u}_0)=\mathbf{0}, \quad k=0,\ldots,c, \\ & \mathbf{h}_k(\mathbf{x}_k,\mathbf{u}_0)\le \mathbf{h}_k^{max}, \quad k=0,\ldots,c, \end{aligned}$$

where f is the objective function, $\mathbf{g}_k$ (respectively $\mathbf{h}_k$) is the set of equality (respectively inequality) constraints for the k-th system configuration ($k=0$ corresponds to the base-case, while $k=1,\ldots,c$ corresponds to the k-th post-contingency state, c being the number of contingencies considered), $\mathbf{h}_k^{max}$ is the system limit, $\mathbf{x}_k$ is the vector of state variables (i.e., complex voltages) for the k-th configuration and $\mathbf{u}_0$ is the vector of base-case control variables. For c contingency scenarios, the problem size of the preventive SCED is roughly $c+1$ times larger than that of the classical (base-case) ED problem. It is worthy noting that solving this problem in a centralized manner for large-scale power systems including numerous contingencies gives rise to prohibitive memory usage and execution times.

In real-world applications, however, many post-contingency constraints are redundant; that is, their absence does not affect the optimal value (Capitanescu et al. 2007). Consequently, a class of algorithms based on contingency filtering techniques, which identify and only add those potentially binding contingencies into the formu-

lation, have been developed (Alsac et al. 1990; Alsac and Stott 1974; Bouffard et al. 2005; Capitanescu et al. 2007; Ernst et al. 2001; Monticelli et al. 1987). For example, the contingency ranking schemes from Bouffard et al. (2005) are achieved by investigating a relaxed preventive SCED problem, where a single contingency along with the base-case is considered one at a time. The ranking methods rely on the information of Lagrangian multipliers or the decrease factor of penalized objective function values, and then select contingencies with a severity index above some threshold for further consideration. Other contingency filtering methodologies (Capitanescu et al. 2007) aim to efficiently identify a minimal subset of contingencies to be added based upon the comparison of post-contingency violations. In addition, an approach using the generalized Benders decomposition to construct the feasibility cut from the Lagrangian multiplier vector of constraints is introduced in Li and McCalley (2009). It shows a significant speedup in terms of computation time. However, the drawback of applying the decomposition technique from convex optimization to the highly non-convex problem is the lack of mathematical convergence analysis. There is no established theoretical guarantee that the algorithm can provide a local minimizer.

The second type of SCED problem is called the corrective SCED, with the underlying assumption that contingency constraint violations can resist up to several minutes without damaging the equipments (Monticelli et al. 1987). The corrective SCED allows post-contingency control variables to be rescheduled, so that it is easier to eliminate violations of contingency constraints than the preventive SCED. The problem can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{x}_0,\ldots,\mathbf{x}_c,\mathbf{u}_0,\ldots,\mathbf{u}_c} \quad & f(\mathbf{x}_0,\mathbf{u}_0) \\ \text{s.t.} \quad & \mathbf{g}_k(\mathbf{x}_k,\mathbf{u}_k)=\mathbf{0}, \quad k=0,\ldots,c, \\ & \mathbf{h}_k(\mathbf{x}_k,\mathbf{u}_k)\le \mathbf{h}_k^{max}, \quad k=0,\ldots,c, \\ & |\mathbf{u}_k-\mathbf{u}_0|\le \Delta\mathbf{u}_k^{max}, \quad k=0,\ldots,c, \end{aligned}$$

where $\mathbf{u}_k$ is the control variables for the k-th contingency, and $\Delta\mathbf{u}_k^{max}$ is the maximal allowed variation of control variables. The optimal cost value of corrective SCED is, in general, smaller than that of preventive SCED, but its solution is often harder to obtain since it introduces additional decision variables and nonlinear constraints. Monticelli et al. (1987) tackle the optimization problem by rewriting it in terms of only the contingency-free state variables and control variables, while constraint reductions are represented as implicit functions of these contingency-free state and control variables, which in turn are related to the infeasibility post-contingency operating subproblems. The solution algorithm then becomes an application of the generalized Benders decomposition (Geoffrion 1972) that iteratively solves a base-case ED and separate contingency analysis. Moreover, an extension of a contingency filtering technique from Capitanescu et al. (2007) is studied in Capitanescu and Wehenkel (2008), which features an additional optimal power flow module to verify the controllability of post-contingency states. In Phan and Kalagnanam (2012), Phan and

Kalagnanam present a new Benders cut that makes use of not only the information of the Lagrange multipliers of the linear constraints but also those of nonlinear ones as well as the bounds on the variables. They show that if the conventional Benders cut is used without care, the generated solution will be rather far from the optimal operating point. By taking the non-convexity into account properly, their adaptive Benders decomposition often improves the quality of the solution. In addition, based on the reformulation of the original problem, they are able to apply the alternating direction method of multipliers (Gabay and Mercier 1976; Glowinski and Marrocco 1975) to decompose the large-scale problem into smaller subproblems.

The third type of SCED problem is an improvement of the aforementioned corrective SCED. Capitanescu and Wehenkel (2007) recognize that the system could face voltage collapse and/or cascading overload right after a contingency and before corrective action is taken. Therefore, their improved formulation imposes existence and viability constraints on the short-term equilibrium reached just after contingency occurrence and before corrective controls are applied:

$$
\begin{aligned}
\min \quad & f(\mathbf{x}_0, \mathbf{u}_0) \\
\text{s.t.} \quad & \mathbf{g}_k(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{0}, \quad k = 0, \ldots, c, \\
& \mathbf{g}_k^0(\mathbf{x}_k^0, \mathbf{u}_k) = \mathbf{0}, \quad k = 0, \ldots, c, \\
& \mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k) \leq \mathbf{h}_k^{max}, \quad k = 0, \ldots, c, \\
& \mathbf{h}_k^0(\mathbf{x}_k^0, \mathbf{u}_k) \leq p_k \mathbf{h}_k^{max}, \quad k = 0, \ldots, c, \\
& |\mathbf{u}_k - \mathbf{u}_0| \leq \Delta \mathbf{u}_k^{max}, \quad k = 0, \ldots, c,
\end{aligned}
$$

where $p_k \geq 1$ is a scalar value modeling how much the constraints just after the contingency application are relaxed with respect to the permanent limits. Additional constraints $\mathbf{g}_k^0$ and $\mathbf{h}_k^0$ are imposed to maintain an appropriate intermediate state between the contingency occurrence and the corrective actions. There are very limited solution approaches to this formulation; but one exception is the work by Li (2008) that utilizes the Benders decomposition method to solve this problem.

The inclusion of contingencies beyond N-1 for future power grid operation may increase the complexity and scale of the problem by several orders of magnitude. Therefore, great effort has been devoted to the development of parallel algorithms for large-scale problem formulations. In this case the SCED problem is decomposed and distributed on a number of processors with each one independently handling a subset of the post-contingency analysis.

There are currently two promising approaches for parallelism: one related to interior-point methods (Qiu et al 2005) and one using Benders decomposition (Alves et al. 2007; Li 2008; Rodrigues et al. 1994). For interior-point methods, at each primal-dual iteration, we need to solve a large-scale system of linear equations. Because the matrix associated with these linear equations has a blocked-diagonal bordered structure, researchers, exploiting this fact, have shown that the system of linear equations can be solved efficiently in parallel. For example, more than ten times speedup can be obtained on a system with sixteen processors (Li and McCalley 2009).

On the other hand, Benders decomposition is a two-stage solution method consisting of a base-case problem and a list of contingency subproblems. Since the evaluation of different contingencies can be done independently, this formulation is amenable for parallelism. One obvious benefit of exploiting parallelism in solving these problems is that it makes the complexity linearly dependent on the size of the problem, as opposed to the quadratic growth in sequential computation (Li 2008). All the above-mentioned algorithms, however, have appeared in the form of academic papers. Their practical use has not been validated.

Another challenge to the existing SCED formulations come from the emergence of deregulated electricity markets, where the market-clearing process, pricing mechanism, and electricity trading should be included as part of the solution process. New formulations and algorithms need to be developed to address this challenge, and the solution should guarantee a satisfactory worst-case performance to meet the real-time dispatching requirements.

7 Concluding Remarks and Further Research Directions

Many of the grid system changes are inherently stochastic in nature, and they cannot be understood through the deterministic approaches currently in use. Future power grid operations should incorporate stochastic modeling of both generation and loading and comprehensive contingency analysis beyond existing N-1 practices in the formulation of both UC and ED problems. Extremely fast algorithms for security-constrained UC and ED with stochastic analysis and integrated real-time N-k contingency analysis will be a critical capability for the evolving smart grid.

Existing power system optimization problems (such as UC and SCED) are mostly based on DC power flow formulations because of the difficulties related to AC power flow computation, including poor convergence rate and non-robust solution quality. However, it is desirable to incorporate AC-based formulation into existing power system optimization problems since the AC-based formulation captures the physical power flow more realistically than the DC-based formulation.

We point out a number of important avenues of research that will receive noteworthy attention in the coming decade.

7.1 Solving Real-Life Instances of UC and ED Problems with Thousands of Generators and Substations in Real-Time

Current practice is that many ISOs use general-purpose integer programming solvers for their UC problems, and these solvers can handle UC problems with number of generators in the order of hundreds and with 72 time periods. But in reality, ISOs need to deal with thousands of generators and probably even more in the future, as

more distributed generation sources come on line. Hence it is very critical to be able to solve large-scale UC and ED problems in real-time. In this respect, approximation algorithms that guarantee finding near-optimal solutions and distributed decomposition algorithms that are scalable will be among the most promising techniques. Similarly, algorithms that lend themselves to parallelization on distributed or shared memory computing environments will gain importance. In addition, it is also crucial to develop algorithms that are robust with fast convergence rates to achieve high-quality solutions to the nonlinear AC-based UC/SCED problems.

7.2 Simultaneously Solving both UC and ED with AC Transmission Constraints

In many real-life practices, linear approximations of transmission constraints are used in the UC problem. Also, there is some research on handling nonlinear transmission constraints of power flow in ED, independent of the UC decisions. Ideally, we would like to solve the UC problem together with the AC-based ED along with the original nonlinear transmission constraints to obtain more realistic results.

The problem can be formulated as a mixed integer nonlinear programming (MINLP) problem, a subclass of both mixed integer programming (MIP) and nonlinear programming problems. The state-of-the-art MIP solvers are now able to handle problem sizes up to hundreds of integer variables efficiently. But MINLP has yet to reach that level of maturity. For the future power system applications, we should exploit the problem structure and leverage the recent advancements in solving large-scale NLP problems so that we can achieve good-quality solutions to MINLP at least as good as MIP. In particular, it will be an important area of research to develop decomposition algorithms that handle transmission constraints and UC decisions separately, and combine them in a way that commitment and transmission decisions optimize a centralized objective function. This will exploit the algorithms developed separately for the UC and AC-based ED problems. Note that the presence of discrete control variables, such as transformer taps, shunt capacitor banks, and other FACTS devices, also makes even the ED problem a MINLP, which will also benefit from above research.

7.3 Security-Constrained UC and ED with Stochastic Analysis

In reality, many problems should be treated stochastically (Hu et al. 2010; Wang and Singh 2006; Schellenberg et al. 2006), which severely complicates the optimization problems. At the very least, it makes the size of any deterministic reformulation much larger. But most of these problems are highly structured. To develop a good solver, one should try to exploit such structures from both optimization and linear algebra per-

spectives. This has been done to a limited extent in Alguacil and Conejo (2000) using generalized Benders decomposition algorithm (Benders 1962; Geoffrion 1972).

Accounting for the contingencies of operations, such as generator and transmission outages, is critical in maintaining a reliable and secure power system. These contingencies can be addressed in two ways: First is to design a robust system with the probability of failure minimized. The second one is to overdesign the system where some excess capacities in the generators and transmission lines are designed so that in the event of an outage, the system can meet the demand using its excess reserves (Gooi et al. 1999). This results in an optimization problem determining how much of the capacity to allocate to reserves so that the probability of failure is kept under control while the total cost of the system operation is minimized.

Countries around the globe have been encouraging and continue to encourage the integration of renewable energy into their power systems. However, renewable energy sources, such as solar and wind power, are highly intermittent and unpredictable. For instance, there are several issues with the wind power: wind speed ramps up and down very quickly leading to fluctuations in the wind power; wind power can be generated only when the wind speed is between a lower and upper limit; and it is very difficult to provide an accurate day-ahead forecast for the wind power. Similar problems, although less severe, exist with solar and other renewable energy sources. Thus, it is very critical to be able to plan the UC and ED decisions considering these uncertainties.

There are two ways to incorporate intermittent renewables into the market. One approach is to assume that all of the renewable energy has to be used to meet the demand. This leads to a UC problem with stochastic demand values, which can be modeled by stochastic programming (Takriti et al. 1996) and robust optimization (Jiang et al. 2011) techniques. In addition, stochastic programming requires building a representative probability distribution for the renewable energy source. Another approach is to buffer against these fluctuations by energy storage. This leads to a problem similar to the production planning problems with inventory decisions.

We believe optimization under uncertainty will be key in addressing the stochastic nature of the power system operation, which results from uncertain fuel prices and load requirements, distributed and intermittent energy generation, evolving demand response, and generation and transmission outages. An efficient UC/SCED tool should be developed that handles real-time multiple-scenario analyses and ensures the consistency of longer forecast horizons with day-ahead markets. These capabilities, which will increase the computational complexity significantly, will become more pronounced when large amounts of renewable sources and demand response programs are integrated into the market causing the loads to be more elastic, dynamic, and uncertain.

In conclusion, unit commitment and economic dispatch are critical for secure power grid operations and one of their main objectives is to maximize market efficiency. It will be necessary to develop a hybrid computing framework and software tools that can utilize new algorithms and mathematics to address these challenges efficiently.

Acknowledgments The authors would like to thank Andrew Conn, Peter Feldmann, Brian Gaucher, Bhavna Agrawal, Jayant Kalagnanam, and Jinjun Xiong for their many valuable comments and suggestions during the course of this work.

References

Acar E, Agrawal B, Conn AR, Ditlow G, Feldmann P, Finkler U, Gaucher B, Gupta A, Heng F, Kalagnanam J, Koc A, Kung D, Phan D, Singhee A, Smith B, Xiong J (2011) Framework for large-scale modeling and simulation of electricity systems for planning, monitoring, and secure operations of next generation electricity grids. In: Proceedings of Computational Needs for the Next Generation Electric Grid, pp 1–73

Acha E, Fuerte-Esquivel CR, Ambriz-Pérez H, Angeles-Camacho C (eds) (2005) FACTS: modelling and simulation in power networks. Wiley, Chichester

Ahmed A, King A, Parija G (2003) A multistage stochastic integer programming approach for capacity expanstion under uncertainty. J Global Optim 26:3–24

Alguacil N, Conejo A (2000) Multiperiod optimal power flow using Benders decomposition. IEEE Trans Power Syst 15(1):196–201

Alsac O, Bright J, Prais M, Stott B (1990) Further developments in LP-based optimal power flow. IEEE Trans Power Syst 5(3):697–711

Alsac O, Stott B (1974) Optimal load flow with steady state security. IEEE Trans Power App Syst PAS-93(3):745–751

Alves JMT, Borges CLT, Filho ALO (2007) Distributed security constrained optimal power flow integrated to a dsm based energy management system for real time power systems security control. In: VECPAR'06: Proceedings of the 7th international conference on high performance computing for computational science, Springer, Berlin, pp 131–144

van Amerongen RAM (1988) Optimal power flow solved with sequential reduced quadratic programming. Electr Eng 71(3):213–219

Aoki A, Satoh T, Itoh M, Ichimori T, Masegi K (1987) Unit commitment in a large scale power system including fuel constrained thermal and pumped storage hydro. IEEE Trans Power Syst 2:1077–1084

Arroyo JM, Conejo AJ (2000) Optimal response of a thermal unit to an electricity spot market. IEEE Trans Power Syst 15:1098–1104

Arroyo JM, Conejo AJ (2004) Modeling of start-up and shut-down power trajectories of thermal units. IEEE Trans Power Syst 19:1562–1568

Ayoub AK, Patton AD (1971) Optimal thermal generating unit commiment. IEEE Trans Power App Syst 90:1752–1756

Bai X, Wei H, Fujsawa K, Wang Y (2008) Semidefinite programming for optimal power flow problems. Int J Electr Power Energy Syst 30:383–392

Baldick R (1995) The generalized unit commitment problem. IEEE Trans Power Syst 10:465–475

Baldwin CJ, Dale KM, Dittrich RF (1959) A study of the economic shutdown of generating units in daily dispatch. AIEEE Trans Power App Syst Part III 78:1272–1284

Baptisella LFB, Geromel JC (1980) Decomposition approach to problem of unit commitment schedule for hydrothermal systems. IEEE Proc Control Theory Appl 127(6):250–258

Baptista EC, Belati EA, da Costa GRM (2005) Logarithmic barrier-augmented Lagrangian function to the optimal power flow problem. Int J Electr Power Energy Syst 27(7):528–532

Bard JF (1988) Short-term scheduling of thermal-electric generators using lagrangian relaxation. Oper Res 36(5):756–766

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. Numer Math 4:238–252

Bertsekas DP, Lauer GS, Sandell NR, Posbergh TA (1983) Optimal short-term scheduling of large-scale power systems. IEEE Trans Autom Control 28:1–11

Bertsimas D, Litvinov E, Sun XA, Zhao J, Zheng T (2013) Adaptive robust optimization for the security constrained unit commitment problem. IEEE Trans Power Syst 28(1):52–63

Birge J, Takriti S (2000) Using integer programming to refine lagrangian-based unit commitment solutions. IEEE Trans Power Syst 15:151–156

Bond SD, Fox B (1986) Optimal thermal unit scheduling using improved dynamic programming algorithm. In: IEEE Proceedings of generation, transmission and distribution, vol 133. pp 1–5

Bouffard F, Galiana FD, Arroyo JM (2005) Umbrella contingencies in security constrained optimal power flow. In: 15th power systems computation conference (PSCC 05). Liège, Belgium

Brini S, Abdallah HH, Ouali A (2009) Economic dispatch for power system included wind and solar thermal energy. Leonardo J Sci 14:204–220

Byrd RH, Nocedal J, Waltz RA (2006) KNITRO: An integrated package for nonlinear optimization. In: di Pillo G, Roma M (eds) Large scale nonlinear optimization. Springer, New York, pp 35–59

Capitanescu F, Glavic M, Ernst D, Wehenkel L (2007a) Contingency filtering techniques for preventive security-constrained optimal power flow. IEEE Trans Power Syst 22(4):1690–1697

Capitanescu F, Glavic M, Ernst D, Wehenkel L (2007b) Interior-point based algorithms for the solution of optimal power flow problems. Electr Power Syst Res 77(5–6):508–517

Capitanescu F, Wehenkel L (2007) Improving the statement of the corrective security-constrained optimal power flow problem. IEEE Trans Power Syst 22(2):887–889

Capitanescu F, Wehenkel L (2008) A new iterative approach to the corrective security-constrained optimal power flow problem. IEEE Trans Power Syst 23(4):1533–1541

Carpentier J (1962) Contribution to the economic dispatch problem. Bull. Soc. Franc. Elect. 8(3):431–447

Carpentier P, Cohen G, Culioli JC, Renaud R (1996) Stochastic optimization of unit commitment: a new decomposition framework. IEEE Trans Power Syst 11:1067–1073

Chang G, Tsai Y, Lai C, Chung J (2004) A practical mixed integer linear programming based approach for unit commitment. In: IEEE PES general meeting, vol 1, pp 221–225

Chen H, Wang X (2002) Cooperative coevolutionary algorithm for unit commitment. IEEE Trans Power Syst 16:128–133

Cheung K, Wang X, Chiu BC, Xiao Y, Rios-Zalapa R (2010) Generation dispatch in a smart grid environment. In: Innovative Smart Grid Technologies (ISGT). Washington, USA, pp 1–6

Chiang HD, Wang B, Jiang QY (2009) Applications of trust-tech methodology in optimal power flow of power systems. In: Kallrath J, Pardalos PM, Rebennack S, Scheidt M (eds) Optimization in the energy industry, energy systems. Springer, Berlin, pp 297–318

Cohen AI, Sherkat VR (1987) Optimization-based methods for operations scheduling. In: Proceedings of the IEEE 75:1574–1591

Cohen AI, Wan AH (1987) A method for solving the fuel constrained unit commitment problem. IEEE Trans Power Syst 2:608–614

Cohen AI, Yoshimura M (1983) A branch-and-bound algorithm for unit commitment. AIEEE Trans Power App Syst Part III 102:444–451

Contaxis GC, Delkis C, Korres G (1986) Decoupled optimal power flow using linear or quadratic programming. IEEE Trans Power Syst PWRS-1:1–7

Dasgupta D, McGregor DR (1994) Thermal unit commitment using genetic algorithms. In: IEE Proceedings, generation, transmission and distribution 141:459–465

Dillon T, Edwin K, Kochs H, Taud R (1978) Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination. IEEE Trans Power App Syst 97:2154–2166

Dragoon K, Milligan M (2003) Assessing wind integration costs with dispatch models: a case study of PacifiCorp. Windpower 2003, Austin

Ernst D, Ruiz-Vega D, Pavella M, Hirsch PM, Sobajic D (2001) A unified approach to transient stability contingency filtering, ranking and assessment. IEEE Trans Power Syst 16(3):435–443

FERC (2006) Security constrained economic dispatch: definition, practices, issues and recommendations—a report to congress regarding the recommendations of regional joint boards for the study of economic dispatch pursuant to Section 223 of the Federal Power Act as added by Section 1298 of the Energy Policy Act of 2005. Technical report, Federal Energy Regulatory Comission, 31 July 2006. See http://www.ferc.gov/industries/electric/indus-act/joint-boards/final-cong-rpt.pdf

Fu W, McCalley JD (2001) Risk based optimal power flow. In: 2001 IEEE Porto Power Tech Conference, Porto, Portugal

Fu Y, Shahidehpour SM, Li Z (2005) Security-constrained unit commitment with AC constraints. IEEE Trans Power Syst 20:1538–1550

Gabay D, Mercier B (1976) A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput Math Appl 2:17–40

Gan D, Thomas R, Zimmerman R (2000) Stability-constrained optimal power flow. IEEE Trans Power Syst 15(2):535–540

Garver LL (1962) Power generation scheduling by integer programming—development of theory. AIEEE Trans Power App Syst Part III 81:730–734

Geoffrion AM (1972) Generalized Benders decomposition. J Optim Theory Appl 10(4):237–260

Ghosh S, Kalagnanam JR, Katz D, Squillante MS, Zhang X (2011) Integration of demand response and renewable resources for power generation management. In: Proceedings of 1st IEEE power engineering society ISGT meeting

Glavitsch H, Bacher R (1991) Optimal power flow algorithms. In: Leondes CT (ed) Analysis and control system techniques for electric power systems, vol 41. Academic Press, New York

Glover JD, Sarma MS, Overbye TJ (2008) Power systems analysis and design. Thomson Learning, Toronto

Glowinski R, Marrocco A (1975) Sur l'approximation par èlèments finis d'ordre un, et la rsolution, par pnalisation-dualit, d'une classe de problèmes de Dirichlet non lineaires. RAIRO Anal Numèr 2:41–76

Gooi H, Mendes D, Bell K, Kirschen D (1999) Optimal scheduling of spinning reserve. IEEE Trans Power Syst 14(4):1485–1492

Gröve-Kuska N, Römisch W (2005) Stochastic unit commitment in hydrothermal power production planning. In: Wallace SW, Ziemba WT (eds) Applications of Stochastic Programming, SIAM, Philadelphia, pp 633–653

Guan X, Luh PB, Amalfi JA (1996) An optimization-based method for unit commitment. Int J Electr Power Energy Syst 14:9–17

Habibollahzadeh H, Bukenko JA (1986) Application of decomposition techniques to short-term operation planning of hydrothermal power system. IEEE Trans Power Syst 1:41–47

Hara K, Kimura M, Honda N (1966) A method for planning economic unit commitment and maintenance of thermal power systems. IEEE Trans Power App Syst 85:427–436

Hatami AR, Seifi H, Sheikh-El-Eslami MK (2009) Hedging risks with interruptible load programs for a load serving entity. Decis Support Syst 48(1):150–157

Hedman K, O'neill R, Oren S (2009) Analyzing valid inequalities of the generation unit commitment problem. In: Power systems conference and exposition, pp 1–6

Hobbs WJ, Hermon G, Warner S, Shelbe GB (1988) An enhanced dynamic programming approach for unit commitment. IEEE Trans Power Syst 3:1201–1205

Hu Z, Wang X, Taylor G (2010) Stochastic optimal reactive power dispatch: Formulation and solution method. Int J Electr Power Energy Syst 32(6):615–621

Huang SJ (2001) Enhancement of hydroelectric generation scheduling using ant colony system based optimization approaches. IEEE Trans Energy Conver 16:296–301

Jabr RA (2003) A primal-dual interior-point method to solve the optimal power flow dispatching problem. Optim Eng 4(4):309–336

Jabr RA (2006) Radial distribution load flow using conic programming. IEEE Trans Power Syst 21(3):1458–1459

Jabr RA (2008) Optimal power flow using an extended conic quadratic formulation. IEEE Trans Power Syst 23(3):1000–1008
Jiang Q, Geng G (2010) A reduced-space interior point method for transient stability constrained optimal power flow. IEEE Trans Power Syst 25(3):1232–1240
Jiang Q, Geng G, Guo C, Cao Y (2010) An efficient implementation of automatic differentiation in interior point optimal power flow. IEEE Trans Power Syst 25(1):147–155
Jiang R, Zhang M, Li G, Guan Y (2011) Benders decomposition for the two-stage security constrained robust unit commitment problem. Optimization. http://www.optimization-online.org/DB_HTML/2011/07/3102.html
Juste KA, Kita H, Tanaka E, Hasegawa J (1999) An evolutionary programming solution to the unit commitment prolem. IEEE Trans Power Syst 14:1452–1459
Kazarlis SA, Bakirtzis AG, Petridis V (1996) A genetic algorithm solution to the unit commitment problem. IEEE Trans Power Syst 11:83–92
Kerr R, Scheidt J, Jr AF, Wiley J (1966) Unit commitment. IEEE Trans Power App Syst 85:471–421
Kim BH, Baldick R (1997) Coarse-grained distributed optimal power flow. IEEE Trans Power Syst 12(2):932–939
Kim BH, Baldick R (2000) A comparison of distributed optimal power flow algorithms. IEEE Trans Power Syst 15(2):599–604
Koc A, Ghosh S (2012) Optimal scenario tree reduction for the stochastic unit commitment problem. In: Winter simulation conference, pp 1–12
Koc A, Kalagnanam J (2012) Parallel branch-cut-price for solving stochastic unit commitment problems for the smart grid (submitted)
Lai SY, Baldick R (1999) Unit commitment with ramp multipliers. IEEE Trans Power Syst 14:58–64
Lam A, Zhang B, Tse D (2011) Distributed algorithms for optimal power flow problem. http://arxiv.org/abs/1109.5229
Lauer GS, Sandell NR, Bertsekas DP, Posbergh TA (1982) Solution of large-scale optimal unit commitment problems. IEEE Trans Power App Syst 101:79–86
Lavaei J (2011) Zero duality gap for classical OPF problem convexifies fundamental nonlinear power problems. In: American control conference, pp 4566–4573
Lavaei J, Low S (2012) Zero duality gap in optimal power flow problem. IEEE Trans Power Syst 27(1):92–107
Lee FN (1991) The application of commitment utilization factor to the thermal unit commitment. IEEE Trans Power Syst 6:691–698
Lee J, Leung J, Margot F (2004) Min-up/min-down polytopes. Discrete Optim 1:77–85
Li C, Johnson RB, Svoboda AJ (1997) A new unit commitment method. IEEE Trans Power Syst 12:113–119
Li H, Li Y, Li Z (2007) A multiperiod energy acquisition model for a distribution company with distributed generation and interruptible load. IEEE Trans Power Syst 22(2):588–596
Li S, Sahidehpour SM, Wang C (1993) Promoting the application of expert system in short-term unit commitment. IEEE Trans Power Syst 3:286–292
Li T, Shahidehpour SM (2005) Price-based unit commitment: a case of Lagrangian relaxation versus mixed integer programming. IEEE Trans Power Syst 20:2015–2025
Li Y (2008) Decision making under uncertainty in power system using benders decomposition. PhD thesis, Iowa State University, Ames, Iowa
Li Y, McCalley JD (2009) Decomposed SCOPF for improving effciency. IEEE Trans Power Syst 24(1):494–495
Liang RH, Kang FC (2000) Thermal generating unit commitment using an extended mean field annealing neural network. In: IEE Proceedings generation, transmission and distribution 147:164–170
Lin WM, Cheng FS, Tsay MT (2002) An improved tabu search for economic dispatch with multiple minima. IEEE Trans Power Syst 17:108–112
Liu C, Shahidehpour SM, Wu L (2010) Extended benders decomposition for two-stage SCUC. IEEE Trans Power Syst 25:1192–1194

Lowery PG (1966) Generating unit commitment by dynamic programming. IEEE Trans Power App Syst 85:422–426
Lulli G, Sen S (2004) A branch-and-price algorithm for multistage stochastic integer programming with application to stochastic batch-sizing problems. Manage Sci 50:786–796
Ma H, Shahidehpour SM (1998) Transmission-constraint unit commitment based on benders decomposition. Int J Electr Power Energy Syst 20:287–294
Maifeld TT, Sheble GB (1996) Genetic-based unit commitment algorithm. IEEE Trans Power Syst 11:1359–1370
Mantawy A, Abdel-Magid Y, Selim S (1998) Unit commitment by tabu search. In: IEE Proceedings, generation, transmission and distribution 145:56–64
Mantawy A, Abdel-Magid Y, Selim S (1999) Integrating genetic algorithms, tabu search and simulated annealing for the unit commitment problem. IEEE Trans Power Syst 14:829–836
Mantawy AH, Soliman SA, El-Hawary ME (2002) A new tabu search algorithm for the long-term hydro scheduling problem. In: Proceedings of large engineering systems conference, power engineering, pp 29–34
Merlin A, Sandrin P (1983) A new method for unit commitment at electricite de france. IEEE Trans Power App Syst 102:1218–1225
Min W, Shengsong L (2005) A trust region interior point algorithm for optimal power flow problems. Int J Electr Power Energy Syst 27(4):293–300
Mokhtari S, Singh J, Wollenberg B (1987) A unit commitment expert system. In: Proceedings of the PICA, pp 400–405
Monticelli A, Pereira MVF, Granville S (1987) Security-constrained optimal power flow with post-contingency corrective rescheduling. IEEE Trans Power Syst 2(1):175–180
Mori H, Matsuzaki O (2001) Embdedding the priority list into tabu search for unit commitment. In: Proceedings of Power Engineering Society Winter Meeting, pp 1067–1072
Muckstadt A, Koenig SA (1977) An application of Lagrangian relaxation to schediling on power generation systems. Oper Res 25(3):387–403
Muckstadt JA, Wilson RC (1968) An application of mixed integer programming duality to scheduling thermal generating systems. IEEE Trans Power Syst 87:1968–1977
Mukerji R (2010) NYISO day-ahead unit commitment design. In: Technical conference on unit commitment software. Federal Energy Regulatory Commission, Washington DC
Nowak MP, Romisch W (2000) Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. Ann Oper Res 100:251–272
Ott A (2010) Unit commitment in PJM. In: Technical conference on unit commitment software. Federal Energy Regulatory Commission, Washington DC
Padhy N (2004) Unit commitment—a bibliographical survey. IEEE Trans Power Syst 19:1196–2005
Padhy NP (2001) Unit commitment using hybrid models: a comparative study for dynamic programming, expert system, fuzzy system, and genetic algorithms. Int J Electr Power Energy Syst 23:827–836
Pang CK, Chen HC (1976) Optimal short-term thermal unit commitment. IEEE Trans Power App Syst 95:1336–1346
Pang CK, Sheble GB, Albuyeh F (1981) Evaluation of dynamic programming based methods and multiple area representation for thermal unit commitments. IEEE Trans Power App Syst 100:1212–1218
Phan DT (2012) Lagrangian duality and branch-and-bound algorithms for optimal power flow. Oper Res 60(2):275–285
Phan DT, Ghosh S (2011) A two-stage non-linear program for optimal electrical grid power balance under uncertainty. In: Proceedings of the 2011 winter simulation conference, pp 4222–4233
Phan DT, Kalagnanam J (2012) Distributed methods for solving the security-constrained optimal power flow problem. In: Proceedings of the 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), pp 1–7

Qiu W, Flueck AJ, Tu F (2005) A new parallel algorithm for security constrained optimal power flow with a nonlinear interior point method. In: IEEE Power Engineering Society General Meeting, pp 2422–2428

Quyang Z, Shahidehpour S (1990) Short-term unit commitment expert system. Electr Power Syst Res 20:1–13

Quyang Z, Shahidehpour SM (1992) An itelligent dynamic programming for unit commitment application. IEEE Trans Power Syst 6:1203–1209

Rajan D, Takriti S (2005) Minimum up/down polytopes of the unit commitment problem with start-up costs. Technical report, IBM Research

Rodrigues M, Saavedra OR, Monticelli A (1994) Asynchronous programming model for the concurrent solution of the security constrained optimal power flow problem. IEEE Trans Power Syst 9(4):2021–2027

Rothleder M (2010) Unit commitment in the CAISO market. In: Technical conference on unit commitment software. Federal Energy Regulatory Commission, Washington DC

Ruzic S, Rajakovic N (1991) A new approach for solving extended unit commitment problem. IEEE Trans Power Syst 6:269–275

Salam S (2007) Unit commitment solution methods. World Acad Sci Eng Technol 35:320–325

Santos AJ, da Costa GRM (1995) Optimal power flow solution by Newton's method applied to an augmented Lagrangian function. In: IEE proceedings, generation, transmission and distribution 142(1):33–36

Sasaki H, Watanabe M, Yokoyama R (1992) A solution method of unit commitment by artificial neural networks. IEEE Trans Power Syst 7:974–981

Schellenberg A, Rosehart W, Aguado J (2006) Cumulant-based stochastic nonlinear programming for variance constrained voltage stability analysis of power systems. IEEE Trans Power Syst 21(2):579–585

Sheble G, Fahd G (1994) Unit commitment—literature synopsis. IEEE Trans Power Syst 9:128–135

Singh KJ, Philpott AB, Wood RK (2009) Dantzig-wolfe decomposition for solving multistage stochastic capacity-planning problems. Oper Res 57(5):1271–1286

Sisworahardjo NS, El-Kaib AA (2002) Unit commitment using ant colony search algorithms. In: Proceedings of large engineering systems conference, power engineering pp 2–6

Siu TK, Nash GA, Shawwash ZK (2001) A practical hydro, dynamic unit commitment and loading model. IEEE Trans Power Syst 16:301–306

Snyder WL Jr, Powell HD Jr, Rayburn JC (1987) Dynamic programming approach to unit commitment. IEEE Trans Power Syst 2:339–347

Sojoudi S, Lavaei J (2012) Network topologies guaranteeing zero duality gap for optimal power flow problem. In: Proceedings of IEEE Power and Energy Society General Meeting, pp 1–7

Sousa A, Torres G (2007) Globally convergent optimal power flow by trust-region interior-point methods. In: Power Tech, 2007 IEEE Lausanne, pp 1386–1391

Sousa AA, Torres GL (2011) Robust optimal power flow solution using trust region and interior-point methods. IEEE Trans Power Syst 26(2):487–499

Street A, Oliveira F, Arroyo JM (2011) Contingency constrained unit commitment with n-k security criterion: a robust optimization approach. IEEE Trans Power Syst 26(3):1581–1590

Sun D, Ashley B, Brewer B, Hughes A, Tinney W (1984) Optimal power flow by newton approach. IEEE Trans Power App Syst PAS-103(10):2864–2880

Swarup KS, Yamashiro S (2002) Unit commitment solution methodology using genetic algorithm. IEEE Trans Power Syst 17:87–91

Takriti S, Birge J, Long E (1996) A stochastic model for the unit commitment problem. IEEE Trans Power Syst 11(3):1497–1508

Takriti S, Birge JR (1996) A stochastic model for the unit commitment problem. IEEE Trans Power Syst 11:1497–1508

Takriti S, Krasenbrink B, Wu LSY (2000) Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. Oper Res 48(2):268–280

Tong SK, Sahidehpour SM, Quyang Z (1991) A heuristic short-term unit commitment. IEEE Trans Power Syst 6:1210–1216
Tong SK, Shahidehpour SM (1990) An innovative approach to generation scheduling in large-scale hydro-thermal power systems with fuel constrained units. IEEE Trans Power Syst 5:665–673
Torres G, Quintana V (1998) An interior-point method for nonlinear optimal power flow using voltage rectangular coordinates. IEEE Trans Power Syst 13(4):1211–1218
Turgeon A (1978) Optimal scheduling of thermal generating units. IEEE Trans Autom Control 23:1000–1005
Viana AMMMG (2003) Metaheuristics for the unit commitment problem: the constraint oriented neighbourhoods search strategy. PhD thesis, Universidade do Porto
Virmani S, Imhof K, Mukhenjee S (1989) Implementation of a Lagrangian relaxation based unit commitment problem. IEEE Trans Power Syst 4:1065–1073
Wächter A, Biegler LT (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math Program 106(1):25–57
Wang C, Shahidehpour SM (1993) Effects of ramp-rate limits on unit commitment and economic dispatch. IEEE Trans Power Syst 8:1341–1350
Wang H, Murillo-Sanchez C, Zimmerman R, Thomas R (2007) On computational issues of market-based optimal power flow. IEEE Trans Power Syst 22(3):1185–1193
Wang L, Singh C (2006) Multi-objective stochastic power dispatch through a modified particle swarm optimization algorithm. In: Proceedings of IEEE swarm intelligence symposium, pp 128–135
Wells DW (1968) Method for economic secure loading of a power systems. In: Proceedings of IEEE vol 115, pp 606–614
Wolkowicz H, Saigal R, Vandenberghe L (eds) (2000) Handbook of semidefinite programming: theory, algorithms, and applications. Kluwer Academic Publishers, Boston
Wood AJ, Wollenberg BF (1996) Power Generation Operation and Control. Wiley, New York
Wu L, Shahidehpour SM (2010) Accelerating the Benders decomposition for network-constrained unit commitment problems. Energy Syst 1:339–376
Wu L, Shahidehpour SM, Li T (2007) Stochastic security-consrtained unit commitment. IEEE Trans Power Syst 22:800–811
Xiong P, Jirutitijaroen P (2011) Stochastic unit commitment using multi-cut decomposition algorithm with partial aggregation. In: IEEE PES general meeting, pp 1–8
Xue Y, Chang L, Meng J (2007) Dispatchable distributed generation network—a new concept to advance dg technologies. In: Power Engineering Society general meeting, 2007, IEEE, pp 1–5
Yang H, Yang P, Huang C (1997) A parallel genetic algorithm approach to solving the unit commitment problem: implementation on the transputer networks. IEEE Trans Power Syst 12:661–668
Yong T, Entriken R, Zhang P (2009) Reserve determination for systems with large wind generation. In: IEEE PES general meeting, pp 1–7
Zhang J, Fuller JD, Elhedhli S (2010) A stochastic programming model for a day-ahead electricity market with real-time resrve shorage pricing. IEEE Trans Power Syst 25:703–713
Zhang M, Guan Y (2009) Two-stage robust unit commitment problem. http://www.optimization-online.org/DB_FILE/2009/10/2427.pdf!
Zhuang F, Galiana FD (1988) Toward a more rigorous and practical unit commitment by lagrangian relaxation. IEEE Trans Power Syst 3:763–773
Zhuang F, Galiana FD (1990) Unit commitment by simulated annealing. IEEE Trans Power Syst 5:311–318

Homeostatic Control and the Smart Grid: Applying Lessons from Biology

Martin Beckerman

Abstract Electric power grids in this country and abroad are undergoing revolutionary changes through the increased integration of electric power generation, delivery and consumption with computation, communications, and cyber security. Emerging out of these activities is a smart grid that includes new technologies ranging from microgrids capable of islanded operation to wind power generation and electric vehicle supply. The success of this massive endeavor will depend on large measure on the development of control methodologies that maintain homeostasis in the face of natural stresses, malfunctions and deliberate attacks. The goal of this chapter is to sketch out possible control strategies for the future smart grid based upon insights into how living systems deal with these same issues. This is a broad topic and the particular focus here will be on presenting a simple model of control by neural and innate immune systems that could be applied to operational security at substations and microgrids.

Keywords Operational security · Innate immunity · Neural control · Multi-agent systems · Substations · Microgrids

1 Introduction

Smart Grids have their beginnings in the pioneering work of MIT professor Fred Schweppe. Writing in the 1970–1980 timeframe he introduced the novel ideas of homeostatic utility control (Schweppe et al. 1980), in which pricing depended on system conditions and is dynamically adjusted, and household appliance load set-points that are adjusted according to system frequency. These notions have their

M. Beckerman (✉)
Information Technology Division,Y-12 National Security Complex,
Oak Ridge, TN 37831, USA
e-mail: beckermanm@y12.doe.gov

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_2,

culmination in the modern conception of smart meters, which along with microgrids capable of islanded operation are key elements of modern smart electric grid projects now underway throughout the world. These developments are driven by declining energy resources and increasing demand, and by the desire to develop green technologies. Smart meters have been now joined by smart appliances, smart houses and the smart grid, and by wind power and electric cars plugged in at homes and elsewhere serving as reservoirs of electrical power.

Electrical grids are subject to a variety of environmental and internal stresses including human error leading to supply-demand load imbalances, frequency drift and, in worse cases, cascading blackouts. Power systems increasingly operate under highly stressed conditions, and as a result blackouts may be triggered by any of a number of forms of instability. Cascading failures may be triggered, for example, by a loss of voltage or frequency stability, by a combination of the two, and by inter-area oscillations (IEEE/CIGRE 2004). Additional stresses and failure modes become possible due to the integration of communications into the grid making increasing likely deliberate attacks by cybercriminals and agents of hostile nation-states.

Valuable insights into weaknesses in the present electric grid are provided by studies carried out during the past ten or fifteen years of their physical, i.e., topological, properties. An important finding from these studies (Sachtjen et al. 2000; Albert et al. 2004) is that the links between nodes in man-made networks such as the power grid are not normally distributed with most of their links centered about some average value. Instead, the distribution of links follows a power law in which most nodes are connected by just a few links and there are a significant number of highly connected nodes. It is then observed that the electric power grid is exceptionally vulnerable to disruptions in its highly connected nodes. These disruptions can trigger a breakdown of an entire network, fragmenting it into isolated patches. In networks such as these, large-scale blackouts are far more common than would be the case were the probability distribution functions to have conventional exponential tails. This situation is exacerbated by a second key feature of these networks. They typically operate close to a critical point; i.e., close to their operational limits (Sachtjen et al. 2000; Carreras et al. 2004). The power system networks are thus doubly susceptible to cascading failures brought on by deliberate attacks (Dobson et al. 2007).

1.1 Need for Operational Security

Operational security—the investigation, mitigation and recovery from stresses, perturbations, and disruptions, both transient and longer lasting—is a major requirement of the smart grid. To achieve operational security, in other words, to maintain homeostasis, the smart grid must be capable of sequentially carry out the following operations:

- Rapidly detect and respond to loss of homeostasis
- Limit local damage when it occurs

- Initiate inter-site communications
- Prevent cascading failures
- Recover and return to normal operations.

Adding to the urgent need for operational security is the growing threat of cyber-attacks. A noteworthy event in this regard was the emergence of the Stuxnet Siemens worm in early 2010, the first publically recognized malware attack of a supervisory control and data acquisition (SCADA) system. Disruptions of this kind have become ever-more likely due of the strong integration of the power system into the Internet and the widespread use of commercial off-the-shelf (COTS) software in the power grid's control systems. The creation of a smart grid with multiple, readily accessible entry points for insertion of malware will only increase the frequency of disruptions of this kind.

1.2 Goals of this Chapter

The goal of this chapter is to apply to the smart grid lessons learned from studying how biological organisms deal with the loss of homeostasis brought on by pathogen attacks, stresses and injury. Recall that in higher organisms such as us there are two types of responses to invasion by pathogens. The innate immune response develops rapidly, and promotes formation of a protective and isolated environment for clearance, repair and recovery of the damaged tissue. The adaptive immune response develops slowly over several days; it involves generation of antibodies and activation of B and T cells to provide long-lasting immunity. The adaptive immune response is unique to vertebrates, while fungi, plants and animals possess a rapid response system to the onset of infection and tissue damage. Our focus in the chapter is on the rapid response, innate immune system.

In humans there are three super-systems—the immune, nervous, and endocrine systems. Contrary to popular thought these systems do not act independently of one another but rather are interconnected and continually communicate and regulate one another. In particular, the nervous system provides central homeostatic feedback control over the actions of cellular agents of the innate immune system such as macrophages and neutrophils. Acting together, the innate immune and nervous systems provide a coordinated response to stress, injury and invasion. We will begin our exploration with a brief overview of how the nervous system provides central control over the innate immune system. We will then discuss the inflammatory response and multi-agent systems, and introduce a simple neuro-immune model for enhanced operational security at substations and microgrids.

2 Central Control by the Nervous System

Central control over innate immunity by the central nervous system is present in all creatures, great and small. For instance, the nematode worm, *Caenorhabditis elegans*, possesses a primitive nervous system containing just 302 neurons. Yet, it

was found recently (Sun et al. 2011) that neurons in *C. elegans* monitor the inflammatory response and limit its effects arising from pathogens and injury. These neurons provide negative feedback control and maintain innate immunity homeostasis. The nervous system in humans is, of course, far more complicated and regulates the innate immune response through multiple hormonal and neural pathways (Sternberg 2006). It fosters communication with the innate immune system by sharing many of the same chemical messengers (Blalock 1989). It monitors the molecular products of pathogen invasion and the presence of cytokines, the main chemical messengers used by the innate immune system. Once activated, the nervous system first potentiates the immune response and then working through negative feedback loops it damps down the immune response and restores homeostasis within the community of innate immune mediators.

The two primary systems of neural regulators of innate immunity in humans are the hypothalamic-pituitary-adrenal (HPA) axis and the vagus nerve. The vagus nerve provides negative feedback regulation of inflammation through sensory afferent signals and efferent feedback (Tracey 2002; Wang et al. 2003). It is a large nerve (cranial nerve X) extending from the brain stem to the colon and containing numerous branches. It sends efferent output to many of the muscles/organs of the body and is involved in control of heart rate, breathing and digestion. The HPA axis is a neuroendocrine regulator of immunity (Sternberg 2006). This regulatory pathway encompasses the hypothalamus located in the brain, the pituitary gland situated below the hypothalamus at the base of the brain, and the adrenal glands located just above the kidneys. It regulates body temperature, energy levels and digestion, and the body's response to stress, injury and trauma through the fight or flight response. These actions are joined by contributions from the sympathetic and peripheral nervous systems that along with the HPA axis and cholinergic pathway (vagus nerve regulate innate immunity at the local, regional and system-wide levels (Sternberg 2006). This system of control is obviously a complex one with many outstanding questions on how it operates remaining to be uncovered. In Sect. 4 of this chapter, we will replace it with a far simpler and better characterized neural control system that could be used to regulate the actions of the electric grid innate immune system.

3 Innate Immunity and the Inflammatory Response

Everyone is familiar with inflammation, a key component of innate immunity. The inflammatory response produces fever and pain, and redness and swelling. A protected local environment is formed that limits spread of the infection, promotes the destruction of the invaders, and hastens the repair of the damaged tissues. Most importantly, the physiological changes in the local environment facilitate the convergence of white blood cells (leukocytes) to the site of the infection. These white blood cells, principally macrophages and neutrophils, are the cellular agents of the innate immune response. In the first phase of the inflammatory response, these agents destroy pathogens, and remove dead and dying cells, damaged support structure, and

cellular debris. In the second, recovery phase, they help restore the tissue to a healthy, fully-functional condition.

Several kinds of cells—neutrophils, monocytes and macrophages, dendritic cells and mast cells—mediate innate immunity. In addition, there is a (non-cellular) complement system that assists the cellular component through the release of molecules that mark extracellular pathogens for destruction and, along with antigen-presenting dendritic cells and macrophages, activates the adaptive immune response. In the remainder of this section, we will consider a simplified system consisting just of macrophages and focus first on what they do (their patterns of behavior) and then on how they do it (sensors and platforms).

3.1 Macrophage Patterns of Behavior

The first features of macrophage behavior worth noting is that both circulating and tissue resident macrophagesare utilized in innate immunity. Examples of tissue resident macrophages are Kupffer cells (liver) and osteoclasts (bone), alveolar macrophages (lung) and microglia (brain). These cells are assisted by mobile, patrolling monocytes (macrophage precursors) and macrophages that converge on infected and injured tissue.

The second key observation is that macrophages are highly plastic and respond to environmental cues by reversibly switching from one phenotype (behavior) to another. They may carry out pathogen clearance and tissue cleanup responses when those actions are necessary, execute a repair and recovery program whenever that is needed, and perform regulatory tasks to prevent excessive responses if that is required (Mosser and Edwards 2008; Beckerman 2009). The third main point, already hinted at above, is that inflammatory responses are tightly controlled, and control over their agents is exerted at the local, regional and systemic levels by the immune and nervous systems.

Overall, there are multiple patterns of macrophage activity and their actions need to be carefully timed and coordinated—damage detection followed by isolation and cleanup followed by tissue restoration. In the smart grid, these would be replaced by the five action stages key to maintaining operational security listed in Sect. 1.1 and presented in the form of a bullet list.

3.2 Responding to Signals of Invasion and Injury

The seminal concept of innate immunity is that pathogens are detected by specialized germline-encoded sensors. This operational model was introduced by Charles Janeway in (1989). The sensors envisioned by Janeway were capable of detecting molecular patterns characteristic of not just one species of pathogen but rather whole classes of them. The pathogen-associated molecular patterns (PAMPs) are microbial

components that are essential to pathogen survival and cannot be easily discarded or disguised.

In a further development of the model, Polly Matzinger (1994) proposed in 1994 that pattern recognition receptors (PRRs) not only respond to microbial PAMPs but also respond to non-microbial signals indicative of trauma and damage. These other signals are termed damage-associated molecular patterns (DAMPs). Under normal conditions, DAMPs are sequestered in cellular compartments away from the sensory apparatus but when the cells are sufficiently stressed they are released into the cytosol where they are detected and initiate inflammatory responses.

In the past ten years a third major development in innate immunity has taken place. Beginning in 2002 (Martinon et al. 2002) it has become ever clearer that macrophages utilize molecular platforms to response to signals of injury and invasion. These platforms, called inflammasomes (Martinon et al. 2002), bring together in one location PAMP and DAMP sensors, interfaces and the downstream initiators of cellular and system responses. Inflammasomes are positioned at strategic locations throughout the cell where they monitor crucial components (organelles) for indications of damage and malfunction. Utilizing a combinatorial code these response platforms generate the correct responses to the variety of signals being received at a given time.

Two observations provide further insight into how biological organisms respond to pathogens and damage. First, indirect detection of injury and invasion takes place. In these situations, the indicators that are being sensed are produced by the processes triggered by the causative agents, not the agents themselves. The products being sensed are produced rapidly, and their detection launches an immediate inflammatory response that limits the damage to a particular region, and begins the healing and restoration of homeostasis. In many instances, the key properties being sensed are inappropriate signaling and control actions. Second, in many cases, not one but two distinct signals are required to elicit an immune response, for instance, a PAMP and a DAMP. This is done to prevent premature and inappropriate immune actions.

4 Central Pattern Generators

Central pattern generators (CPGs) are elementary circuits built from a small number of neurons that generate the highly stable motor patterns responsible for activities such as walking, swimming, breathing, chewing and digestion (Beckerman 2005). Generation of these behaviors is autonomous—the circuits are self-contained, able to generate rhythmic patterns independent of timing input and sensory feedback. They are not only relatively simple but also readily accessible to experimental manipulation and produce easy to distinguish motor patterns (behaviors). CPGs were first studied a hundred years ago (Graham-Brown 1911) and in the ensuing time period have been the subject of numerous studies, experimental and theoretical, in species ranging from primitive invertebrates to mammals. They have already entered the engineering arena through their use in robotic movement control (Ijspeert 2008).

Several properties of these control circuits are desirable from perspective of the electric grid of the future. In this chapter, we propose that controllers with similar properties be adapted for use in the electric grid, not to generate rhythmic (motor) patterns but rather to work together with immunological agents to generate and coordinate operational security actions and endow the grid with resilience to malfunctions and deliberate attacks. In the remainder of this section, two properties will be examined. The first of these is stability, that is, the ability of these circuits to maintain constancy of output in the face of variability and breakdowns in components. The second is responsiveness, namely, the ability of these circuits to switch from behavioral state to another is response to modulatory signals from outside.

4.1 Redundancy and Degeneracy

To understand how these circuits maintain robustness in the presence of considerable variability in failure modes we have to distinguish between the terms "degeneracy" and "redundancy". The latter term, redundancy, is a familiar one. It refers to the performance of a particularly task by multiple, structurally identical copies of some element(s). It is a traditional way of dealing with circuits and other structures build from fault- and failure-prone elements. Degeneracy differs from this in a fundamental way. In the case of degeneracy, a particular function is carried out by structurally different elements (Edelman and Gally 2001). That is, there exists multiple, non-identical ways to produce a desired output and in each instance one of these solutions is selected.

Several observations made during the past few years support the notion that biological systems extensively exploit degeneracy and not redundancy to achieve robustness and resilience to malfunctions and loss of components. The first set of observations supporting this model is that there is considerable variability in components, in the numbers and biophysical properties of their ion channels, from one neuron to another of a given type (Marder and Goaillard 2006). Furthermore, the circuits built from these component neurons are remarkably stable, in spite of considerable variability in strengths from synapse to synapse across the neurons. It appears that failures in some components are compensated for by adjustments in the response properties of other components such that the overall behavior of the neuron or circuit is preserved (Prinz et al. 2004).

4.2 Stability and Responsiveness

The second key property of these circuits—responsiveness—is built upon their ability to generate not just one but rather several different patterns, or behaviors, in their downstream targets. That is, they are capable of reconfiguring themselves, some components (neurons) becoming more strongly coupled in to the active circuits while

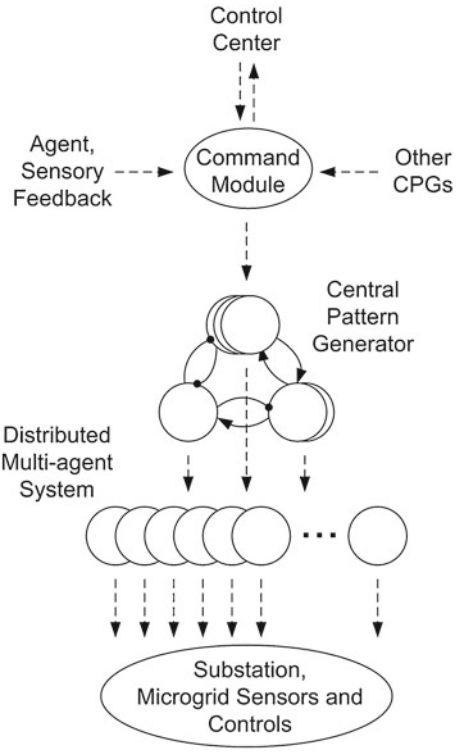

Fig. 1 Biologically-inspired control architecture for the smart grid. Signals supplied through the command module instruct the central pattern generator (CPG) on which pattern of activity to generate among the software agents. The software agents interact with SCADA sensors and controls in substations and their counterparts in microgrids, and supply feedback to the upstream controls. In this model, humans remain in the loop and can direct actions taken. In the CPG circuit, sharp arrows denote excitatory connections and filled circles represent inhibitory connections

others drop out entirely. By this means they can switch from one behavioral state to another, generating several different patterns, each pattern corresponding to a specific functional state on the part of the motor plant they control.

Most significantly, while each of the states is stable against malfunctions, the circuits are responsive to information coming in from outside the circuit. As shown in Fig. 1 the central pattern generator receives input from a modulator (command) module. This module integrates together a variety of signals that jointly guides the selection by the CPG of which pattern to generate. Included in the decision process are signals from other CPGs, from sensory feedback, from feed-forward signals conveying information about external conditions from central control, and from other modulatory signals which for our purposes can be regarded as conveying local information. Thus, these remarkable circuits manage to balance flexibility needed to respond to environmental changes, and to adapt to new demands, with stability against variations in underlying component composition and properties. These are properties needed in the smart grid.

5 Neuro-Immune Model of Homeostatic Control for the Smart Grid

Software agents appear to be an ideal way to incorporate macrophage-like inflammatory responses and innate immunity principles into operational security. This association is enhanced by the close match between the observed patterns of macrophage-inflammasome behaviors and the "bullet list" of necessary operational security actions. Multi-agent systems (MAS), i.e., systems composed of two or more software agents, have been proposed for some time as a means of achieving distributed control over macrogrids in a future smart grid (see (McArthur et al. 2007a, b) for a review of early MAS applications to electric power grid). The software agents in a multi-agent system are encapsulated computer programs capable of flexible, autonomous actions. They are (i) problem solving entities with well-defined boundaries and interfaces; (ii) embedded in specific environments, receiving sensory information about the state of their environment and acting back on the environment through effectors, and (iii) designed to act reactively and proactively in performing their assigned tasks (Wooldridge 1997; Jennings 2001).

As a first step in implementing biologically-inspired operational security, 'inflammasomes' would be positioned where they could monitor and control the operation of key components at substations and microgrids. Signals from these components would be monitored and integrated together at the inflammasome platforms together with signals and data from upstream CPGs in order to assess operating conditions and respond to inappropriate signals and abnormal conditions.

Overall, the combined neuro-immune system serves as an attractive model for the grid. A set of semi-autonomous distributed agents, macrophage-inflammasome (MI) agents, provides for local regulation of the grid with rapid communication and control functions carried out from control centers via neural-like functions. This architecture enables optimization of the power system according to environmental conditions and enables repaid adjustments to the buildup of stresses at the local and regional levels. It enables a rapid dissemination of signals and foster inter-site communications, limiting cascading effects and coordinates rapid responses to alterations in normal operation. Hopefully, it would promote load shedding and microgrid islanding in a way that prevents blackouts that extend over large regions of the future smart grid. These concepts will be discussed in more detail in the remainder of this section.

5.1 High-Level Control Architecture

In the control architecture depicted in Fig. 1, data and alerts are relayed to the control centers via the command module. This information stream (1) informs operators in the control center as to which set of MAS actions are taking place at any given time, and (2) alerts operators in the control centers that a situation has occurred that may require human intervention and repair activities. (3) In new facilities that allow for complete integration of the neural control unit, it may assume many of the duties of

Table 1 Operational security activities by multi-agent systems

Operational step	Agent Behavior	Reference
Substations		
Rapidly detect and respond	Defense	Li et al. (2005)
	Diagnostic support	Buse et al. (2003), Davidson et al. (2006)
	Secondary control	Wang et al. (2003)
Isolate local damage	Islanding, load shedding	M. Pipattanasomporn et al. (2009)
Restore normal operation	Power system restoration	Nagata and Sasaki (2002)
Microgrids		
Rapidly detect and respond	Primary control	Dimeas and Hatziargyriou (2005)
	Diagnostic support	J. Oyarzabal et al. (2005)
	Secondary control	Jimeno et al. (2011)

a firewall. In other, older facilities, it could provide an independent data stream from agents active in the substations and microgrids.

Listed in Table 1 is a partial list of agent behaviors that have been developed by the research groups listed in Column 3. These serve as exemplars of how these control modules would operate in practice at substations and microgrids. These have arranged according to which operational security step they support. In marrying MAS technologies with a neural control module, the basic idea adopted from biology is to provide for central control and coordination by the command module while leaving considerable local autonomy for the agents to cooperate with one another and act according to what they encounter locally. Non-local environmental information that is relevant to their duties is relayed to them in a timely manner via the neural command module. In this biological paradigm, control is spread across neuro-immune levels in a manner that enables each component to optimally contribute to the overall task of maintaining the *milieu intérieur*.

5.2 Substation Control

There are a large number of networked sensing devices and access points embedded within SCADA systems that can be exploited for cyber-attacks (Ericsson 2010; Wei et al. 2011) and should be monitored. These include, for example, protection relays, digital fault recorders (DFRs), LAN switches, remote terminal units (RTUs) and human-machine interfaces (HMIs). Cyber attacks may take one of several forms. They may, for example, be triggered by injection of false data into the substation data stream (Y. Liu et al. 2011) leading to erroneous actions by operators in control centers. The attacks may take a loss or denial of service form if equipment is either turned off or blocked, or alternatively, the attacks may generate physical damage to equipment (as exemplified by Stuxnet).

The first stage in responding to loss of power system homeostasis, whether it is due to supply/demand imbalances, component damage, or deliberate attack, is to

detect the danger and launch an appropriate response. The time scales involved in achieving real-time responsiveness are quite short. Typical response times required within substations are on the order of 100 ms for fault disconnecting and on the order of 100 ms to 2 sec for automation control and monitoring (Ericsson 2010; Wei et al. 2011). A number of pioneering studies on how to achieve these challenging goals have been reported. For example, reference Li et al. (2005) addressed how to protect power system against malicious attacks, presenting a system that might be used to pinpoint dangerous situations in advance of a potentially catastrophic outage. Reference Buse et al. (2003) developed a multi-agent system that managed the large number of data acquisition, monitoring and control systems present in a substation, and reference Davidson et al. (2006) created a multi-agent system that provided diagnostic support and operated in conjunction with SCADA systems and fault recorders. Secondary control refers to maintenance of voltage and frequency stability between control areas. In reference Wang et al. (2003), a multi-agent system for secondary voltage control was explored. As noted earlier the transition from grid-connected to islanded operation is an important component of the smart grid. In reference M. Pipattanasomporn et al. (2009), the facilitation of this transition using MAS technologies was illustrated and reference Nagata and Sasaki (2002) introduced a set of bus agents plus a management agent that assisted in restoration of power in a local network.

5.3 Microgrid Control

An essential feature of the smart grid is the integration of distributed energy resources (DERs) into the low and medium voltage networks. Microgrids are an attractive model for aggregating and integrating some classes of DERs into the grid. They combine DERs such as micro-turbines, wind turbines, and fuel cells with energy storage devices (ESDs) such as batteries and fly-wheels. The DERs and ESDs are then jointly controlled and connected to the low voltage network in ways that enables them to operate in either grid-connected or islanded modes. In assessing the operational security requirements for these entities, it becomes clear that many of the same operational security and control capabilities as used in substations are required. In addition, the control system must handle not only islanding, load shedding and black startup but also market pricing (Dimeas and Hatziargyriou 2005; Lopes et al. 2006).

The utility of agents in operational security has been emphasized in several microgrid studies (J. Oyarzabal et al. 2005; Jimeno et al. 2011). In reference J. Oyarzabal et al. (2005) the utility of agent systems to assume responsibilities carried out by SCADA and other control devices in substations was advanced, while reference Jimeno et al. (2011) explored secondary control applications. Given the increased security risks associated with microgrids, safe and secure integration into the main grid is a challenging task. Making upstream use of a neural command module and pattern generator similar to that advocated for use in substations could facilitate this activity.

6 Conclusion

In this chapter, we sketched how the smart grid of the future might not only guard against natural disasters and deliberate attacks, but also limit damage and promote recovery when these events do occur. The roadmap we presented was modeled on the innate immune system, a highly successful system of protection implemented across multiple phyla and kingdoms. As is the case for biological systems we included in the overall architecture central (neural) control that works alongside the distributed system of innate immune sensor/effector platforms.

With regard to this last point, biological organisms utilize a single, highly integrated immune/neural response and control system to deal with stresses, injury and invasion. The utilization of a single system for dealing with these seemingly different dangers is driven by the fact that many of the same operational response steps are involved in treating each of them. This commonality is strongly reflected by the use of the term sterile inflammation to describe inflammatory responses that occur in damaged tissue for which there is no sign of an invasion (Chen and Nuñez 2010).

The neuro-immune model of electric grid control presented in this chapter is both simplified and abstracted from its biological progenitor. This is done in part to relieve the structure of details tied to wet chemistry that are not relevant in engineering applications. Similarly, details can be expected to enter into developing smart grid descendent that are unique to the grid. One set of "details" may well carry over from biology to engineering is the danger of excessive responsiveness by macrophage/smart grid agents. This aspect was emphasized in this chapter, and several mechanisms to prevent this from happening were discussed.

A number of topics relevant to smart grid resiliency and self-healing were not discussed in this chapter. One of these is the best way to utilize the several means of communication available—optic fiber and the Internet, broadband over power lines (BPL), digital wireless, and microwaves. In biological systems, neural means are favored when rapid long-distance communication is required. In the smart grid, the availability of the aforementioned routes should help in the establishment of an effective analog to neural-like rapid communications.

It is just in the last few years that degeneracy has emerged as a means by which neural systems achieve robust performance. This striking finding is joined by earlier discovered ways CPGs simultaneously achieve stability to small perturbations and responsiveness to environmental changes. These biological properties are clearly desirable ones for the smart grid and will help it achieve resiliency and self-healing. These latter efforts are in their earliest stages, and a large amount of work remains to be done, especially in furthering cyber security against attacks from cyber criminals and agents of hostile nation-states.

References

Albert R, Albert I and Nakarado GL (2004) Structural vulnerability of the North American power grid. Phys Rev E 69:025103(R). doi:10.1103/PhysRevE.69.025103

Beckerman M (2005) Molecular and cellular signaling. Springer, New York

Beckerman M (2009) Cellular signaling in health and disease. Springer, New York

Blalock JE (1989) A molecular-basis for bidirectional communication between the immune and neuro-endocrine systems. Physiol Rev 69:1–32

Buse DP, Sun P, Wu QH, Fitch J (2003) Agent-based substation automation. IEEE Power Energy Mag 50–55. doi:10.1109/MPAE.2003.1192026

Carreras BA, Newman DE, Dobson I, Poole AB (2004) Evidence for self-organized criticality in a time series of electric power system blackouts. IEEE Trans Circuits Syst I 51:1733–1740. doi:10.1109/TCSI.2004.834513

Chen GY, Nuñez G (2010) Sterile inflammation: sensing and reacting to damage. Nat Rev Immunol 10:826–837. doi:10.1038/nri2873

Davidson EM, McArthur SDJ, McDonald JR, Cumming T, Watt I (2006) Applying multi-agent system technology in practice: automated management and analysis of SCADA and digital fault recorder data. IEEE Trans Power Syst 21:559–567. doi:10.1109/TPWRS.2006.873109

Dimeas AL, Hatziargyriou ND (2005) Operation of a multiagent system for microgrid control. IEEE Trans Power Syst 20:1447–1455. doi:10.1109/TPWRS.2005.852060

Dobson I, Carreras BJ, Lynch VE, Newman DE (2007) Complex systems analysis of series of blackouts: Cascading failure, critical points and self-organization. Chaos 17(026103):1–13. doi:10.1063/1.2737822

Edelman GM, Gally JA (2001) Degeneracy and complexity in biological systems. Proc Nat Acad Sci USA 98:13763–13768. doi:10.1073/pnas.231499798

Ericsson GN (2010) Cyber security and power system communication—essential parts of a smart grid infrastructure. IEEE Trans Power Delivery 25:1501–1507. doi:10.1109/TPWRD.2010.2046654

Graham-Brown T (1911) The intrinsic factors in the act of progression in the mammal. Proc R Soc Lond B 84:308–319

IEEE Trans Power Syst (2004) Definition and classification of power system stability. 19:1387–1401. doi:10.1109/TPWRS.2004.825981

Ijspeert AJ (2008) Central pattern generators for locomotion control in animals and robots. Neural Networks 21:642–653. doi:10.1016/j.neunet.2008.03.014

Janeway CA Jr (1989) Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harbor Symp Quant Biol 54:1–13

Jennings NR (2001) An agent-based approach for building complex software systems. Commun ACM 44:35–41. doi:10.1145/367211.367250

Jimeno J, Anduaga J, Oyarzabal J, de Muro AG (2011) Architecture of a microgrid energy management system. European Trans Electrical Power 21:1142–1158. doi:0.1002/etep.443

Li H, Rosenwald GW, Jung J, Liu C (2005) Strategic power infrastructure defense. Proc IEEE 93:918–933. doi:10.1109/JPROC.2005.847260

Liu Y, Ning P, Reiter MK (2011) False data injection attacks against state estimation in electric power grids. ACM Trans Info Syst Sec (TISSEC) 14:Art 13. doi:10.1145/1952982.1952995

Lopes JAP, Moreira CL, Madureira AG (2006) Defining control strategies for microgrids islanded operations. IEEE Trans Power Syst 21:916–924. doi:10.1109/TPWRS.2006.873018

Marder E, Goaillard JM (2006) Variability, compensation and homeostasis in neuron and network function. Nat Rev Neurosci 7:563–574. doi:10.1038/nrn1949

Martinon F, Burns K, Tschopp J (2002) The Inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-β. Mol Cell 10:417–426. doi:10.1016/S1097-2765(02)0059-3

Matzinger P (1994) Tolerance, danger, and the extended family. Annu Rev Immunol 12:991–1045. doi:10.1146/annurev.iy.12.040194.005015

McArthur SDJ, Davidson EM, Catterson VM, Dimeas AL, Hatziargyriou ND, Ponci F, Funabashi T (2007a) Multi-agent systems for power engineering applications—Part 1: concepts, approaches, and technical challenges. IEEE Trans Power Systems 22:1743–1752. doi:10.1109/TPWRS.2007.908471

McArthur SDJ, Davidson EM, Catterson VM, Dimeas AL, Hatziargyriou ND, Ponci F, Funabashi T (2007b) Multi-agent systems for power engineering applications—Part 2: technologies, standards, and tools for building multi-agent systems. IEEE Trans Power Systems 22:1753–1759. doi:10.1109/TPWRS.2007.908472

Mosser DM, Edwards JP (2008) Exploring the full spectrum of macrophage activation. Nat Rev Immunol 8:958–969. doi:10.1038/nri2448

Nagata T, Sasaki H (2002) A multi-agent approach to power system restoration. IEEE Trans Power Syst 17:457–462. doi:TPWRS.2002.1007918

Oyarzabal J, Jimeno J, Ruela J, Engler A and Hardt C (2005) Agent based microgrid management system. In: IEEE conference future power system, pp. 6–11. doi:10.1109/FPS.2005.204230

Pipattanasomporn M, Feroze H and Rahman S (2009) Multi-agent systems in a distributed smart grid: Design and implementation. In: Proceedings of IEEE PES power system conference and exposition (PSCE'09), pp. 1–8. doi:10.1109/PSCE.2009.4840087

Prinz AA, Bucher D, Marder E (2004) Similar network properties from disparate circuit parameters. Nat Neurosci 7:1345–1352. doi:10.1038/nn1352

Sachtjen ML, Carreras BA, Lynch VE (2000) Disturbances in a power transmission system. Phys Rev E 61:4877–4882. doi:10.1103/PhysRevE.61.4877

Schweppe FC, Tabors RD, Kirtley JL, Outhred HR, Pickel FH and Cox AJ (1980) Homeostatic utility control. IEEE Trans Power Apparatus Syst PAS-99:1151–1163. doi:10.1109/TPAS.1980.319745

Sternberg EM (2006) Neural regulation of innate immunity: a coordinated nonspecific host response to pathogens. Nat Rev Immunol 6:318–328. doi:10.1038/nri1810

Sun JR, Singh V, Kajino-Sakamoto R, Aballay A (2011) Neuronal GPCR controls innate immunity by regulating noncanonical unfolded protein response genes. Science 332:729–732. doi:10.1126/Science.1203411

Tracey KJ (2002) The inflammatory reflex. Nature 420:853–859. doi:10.1038/nature01321

Wang HF, Li H, Chen H (2003) Coordinated secondary voltage control to eliminate voltage violations in power system contingencies. IEEE Trans Power Syst 18:588–595. doi:10.1109/TPWRS.2003.810896

Wang H, Yu M, Ochani M, Amella CA, Tanovic M et al (2003) Nicotine acetylcholine receptor $\alpha 7$ subunit is an essential regulator of inflammation. Nature 421:384–388. doi:10.1038/nature01339

Wei D, Lu Y, Jafari M, Skare PM, Rohde K (2011) Protecting smart grid automation systems against cyberattacks. IEEE Trans Smart Grid 2:782–795. doi:10.1109/TSG.2011.2159999

Wooldridge M (1997) Agent-based software engineering. IEE Proc Softw Eng 144:26–37. doi:10.1049/ip-sen:19971026

Operator's Interruption-Cost-Based Sectionalization Method For 3-Feeder Radial Distribution Architecture

Virginijus Radziukynas, Neringa Radziukynienė, Arturas Klementavičius and Darius Naujokaitis

Abstract The distribution system sectionalization (recloser placement) and reconfiguration (closed/open status of reclosers) are receiving increasing interest in the field of feeder design and operation. They are becoming an intrinsic feature of the emerging smart grids by contributing to their reliability and self-healing capability. The majority of sectionalization methods are combinatorial suggesting the development of multiple sectionalization scenarios, and iterative. The goal of the chapter is to present a novel sectionalization method giving an optimal recloser placement scenario at one straightforward calculation. The sectionalization rationale is to split a radial grid into sections with equal portions of operator's interruption cost in the initial configuration, where the interruption cost consists of non-distributed energy cost and fault clearing cost. The optimization is referred to as minimization of non-distributed energy cost for 3-overhead-feeder distribution grid with 3 independent supply points. It provides both single and double sectionalization, with one and two reclosers per feeder, respectively. The method also suggests the evaluation of investment efficiency of the sectionalization in the considered period. It was validated as an effective procedure for the 10-kV test grid in numerical setup with 43 nodes, corresponding to the Lithuanian critical line data. The method-based sectionalization was compared with the reason-based one in order to quantify the optimisation effect.

Keywords Distribution grid · Feeder · Recloser · Sectionalization · Reconfiguration · Optimisation · Non-distributed energy cost

V. Radziukynas (✉) · N. Radziukynienė · A. Klementavičius · D. Naujokaitis
Laboratory of systems control and automation, Lithuanian Energy Institute,
Breslaujos street 3, LT-44403 Kaunas, Lithuania
e-mail: virginijus@mail.lei.lt

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*,
Energy Systems, DOI: 10.1007/978-3-642-38134-8_3,

1 Introduction

The chapter is dedicated to the optimal sectionalization of smart distribution architecture covering three radial overhead feeders with joint terminal intersection point. Its goal is to introduce and validate a novel method of sectionalization which yields a minimum (optimal) non-distribution energy cost and thus brings down the grid operator's interruption cost. The method lends a straightforward procedure to find the optimal sectionalization points and excludes difficulties and troubles specific to combinatorial search and iterative solution process.

The sectionalization of distribution networks is a known means of distributed automation providing effective fault isolation and subsequent reconfiguration. Technically, it is based on reclosers and sectionalizers. Recently, in response to new challenges for modernization of distribution grids, the operators have found new opportunities to benefit from distributed automation, including its sectionalization and reconfiguration functions. The modernization is aimed at upgrading the existing grids to smart grids. The latter grids feature the following seven principal attributes: (1) self-healing, (2) motivating and including customers into grid operation, (3) attack-resistant, (4) providing power quality for the twenty-first century needs, (5) accommodating all generation and storage options, (6) enabling the market and (7) providing asset operation and efficient operation (National Energy Technology Laboratory 2007).

Self-healing is referred to the ability of the grid to perform continuous self-assessments to detect, analyze, respond to, and, if needed, restore its components or sections. Self-healing helps maintain reliability, security, affordability, power quality and efficiency of the grid. The role of reclosers in self-healing is the provision of auto-restoration processes in which distribution circuits employ new protection, communications and control elements able to sense circuit parameters, isolate faults, and quickly, automatically restore service by employing such tools as feeder ties and distributed resources (National Energy Technology Laboratory 2007).

The current enclosers and sectionalizers can often be retrofitted with standardized communications, data processing, and actuating devices to make them compatible with the self-healing requirements of the smart grid. Further, self-healing infrastructure will consist of the transformation of the distribution system from a radial design to an intelligent network design through the addition of circuit-to-circuit ties, the integration of distributed energy resources and demand response, and the application of advanced communication technology (National Energy Technology Laboratory 2010).

Regarding the best practice examples, high-scale feeder automation in Naperville (Illinois, USA) demonstrated a good response to the out-of-power accident. The respective distributed automation project provided 70 % of distribution network with automation. After the outage of 138-kV substation, 25 % of Naperville customers, i.e. approx. 13,600, lost power instantly. In less than 2 min, electricity service was automatically restored for 9,100 customers (67 %), 5,845 thereof were re-energized by enclosers in remote feeder sections, while the rest 3,255 customers—by automatic

load transfer at substation level. Later on, 4,380 customers (32 %) regained service in the next 14 min by means of dispatcher control actions to operate substation bus ties and distribution assets (via SCADA). The remaining 143 customers (1 %) were re-energized by repair crews within 81 min after the accident. Hence, the automated feeders drastically reduced the number of customers exposed to a typical outage. This accident has completely validated the significant investment into the project (Geynisman and Schaub 2007).

Traditionally, a distinction between a circuit breaker and a recloser (autorecloser) is made referring to the incapability of the circuit breaker to reclose: the transient fault would open the breaker or blow the fuse, de-energizing the feeder until a repair crew could manually reclose the circuit breaker or replace the blown fuse. Currently the distinction is not so clear-cut. A circuit breaker is referred to as a self-controlled device for automatically interrupting and reclosing an AC circuit with a preset sequence of openings and reclosures. Circuit breakers do not have built-in controls, thus they open and reclose on the basis of relay operation. A recloser also automatically trips and recloses a preset number of times to clear transient faults or isolate permanent faults. Such unit has a built-in relay and a circuit breaker (Crow and Shetty 2004).

As regards the sectionalizer, it is a self-opening switch that is used in conjunction with the source-side recloser or circuit breaker to automatically isolate faults. When the source-side device operates to de-energize the circuit, the sectionalizer counts the number of operations. After the preset number of operations, the sectionalizer permanently opens, thus limiting the outage only to the faulted section. Another advantage is that it does not have time–current curve and, therefore does not need any coordination Crow and Shetty (2004).

Although the idea of the sectionalization has been worked on for decades, new sectionalization methods and approaches are further developed and presented by addressing different search procedures and rationales (as criteria), different distribution architectures as well as distribution automation technologies. In this case the simplicity and effectiveness of the method (approach) are its crucial characteristics which determine the value of the method.

2 State-of-the-art Research of Feeder Sectionalization and Reconfiguration

Conventional logic suggests placing a recloser at a halfway point of a radial feeder with uniformly distributed load, which, in theory, would yield a 25 % feeder-wide reliability improvement. Similarly, two enclosers should be located at one-third and two-third of feeder length. In reality, due to the presence of critical loads and non-uniform load distributions, utilities often resort to reason-based judgment about the place of enclosers (Greatbanks et al. 2003).

The following review of current approaches and methods of the distribution feeder sectionalization and reconfiguration is presented by distinguishing four major problem classes (situations), namely:

- Reliability-aimed sectionalization;
- Sectionalization and reconfiguration aimed at power losses reduction;
- Sectionalization aspects in optimal allocation of distributed generation;
- Sectionalization and reconfiguration of grid with allocated distributed generation.

Hereafter in the Sect. 2, we adhere to the terminology of references, with variation of synonym terms (as recloser and autorecloser). Also the notion of switch is assumed as a cover term for disconnecting devices relevant for sectionalization and reconfiguration.

2.1 Reliability-Aimed Sectionalization

In pursuance of better reliability, the same authors analyse optimal placement of enclosers and sectionalizers in a 1-feeder grid structure in Refs. (Jamali and Shateri 2005a,b). They define optimality as economic gain over a review period from savings of non-distributed energy (NDE) cost balanced against investment-to-sectionalization cost. The NDE originates from permanent faults. There is no backup independent supply point for grid architecture in question. The recloser is subjected to placement in the sub-transmission substation, while sectionalizers (1, 2 and multiple) are placed along the feeder. The proposed algorithm uses an effective node numbering method to determine the topology of the grid. Fault occurrence probabilities are used in calculus of the NDE cost. The optimal placement is simply taken from the set of all placement scenarios as that with the biggest economic gain. The validity of the method was evaluated by applying the algorithm to a typical 20-kV Iranian feeder.

Similarly to Refs. (Jamali and Shateri 2005a,b), the methodology of autorecloser placement in a 1-feeder grid is aimed at the determination of the most economic location of autoreclosers and their numbers (Hashim et al. 2006). Here the cost of outage for customers is addressed, which formally is a different cost category as compared to non-distributed energy cost. The cost of outage for customers rests upon an estimated Value of Lost Load to quantify the financial losses that customers suffer given a loss of supply. The method does not assume the requirement for cost-effective sectionalization. The cost of outage is summed with the cost of autoreclosers for a particular autorecloser placement scenario to form an objective function. The best placement and numbers of autoreclosers can be decided by determining the least cost option from the said objective function. The same fault probability was prescribed to any segment of a feeder. Test case of a feeder with realistic data of critical Malaysian lines was analysed to check the validity of the methodology.

2.2 Sectionalization and Reconfiguration Aimed at Power Losses Reduction

The sectionalization and resulting reconfiguration possibilities are often considered as a non-reliability problem. The study (Rugthaicharoencheep and Sirisumrannukul 2010) presents an approach for optimal implementation of feeder reconfiguration in the unbalanced loading distribution systems with the objective of power loss reduction. The grid architecture consists of one feeder connected to one independent supply point and split into seven branches (laterals) linked by five ties. The total number of nodes amounts to 69. The optimization problem is solved given the chosen location of sectionalizing switches. Optimization is subjected to system constraints consisting of load-point voltage limits, radial configuration format, absence of load-point interruption, and feeder capability limits. The system power losses and bus voltages are solved by a three-phase power flow algorithm. The solution technique, developed on the basis of Tabu search, is employed to search switch statuses for feeder reconfiguration under different unbalanced loading conditions. The results of the study show that the optimal on/off patterns of the switches which give the minimum power loss can be identified.

The paper (Wu et al. 2007) presents the same problem as in Rugthaicharoencheep and Sirisumrannukul (2010), but for different grid architecture, resting upon different optimisation method. The objective of feeder reconfiguration during normal operation is to find the best switch operation plan for reducing the power losses. The grid consists of five feeders with five independent supply points. The system has 10 sectionalizing switches, 5 tie switches and 15 load-zones. By changing on/off status of sectionalizing and tie switches, grid reconfiguration can be achieved. Since the reconfiguration is done by changing the status of switches, it can be categorized as a discrete combinational optimization problem. Particle Swarm Optimization (PSO) is one of the methods that can be used to solve optimal problems. Typical PSO is designed for continuous function optimization, thus it is not suitable for discrete function optimization. Therefore, the operators of PSO algorithm must be reviewed and redefined to fit the application field of distribution feeder reconfiguration. The paper proposes a method which modifies the operators of PSO formula based on the characters of both status of switches and shift operator to construct a binary coding particle swarm optimization for feeder reconfiguration. The test results show that the proposed method can be effectively applied to feeder reconfiguration.

2.3 Sectionalization Aspects in Optimal Allocation of Distributed Generation

This class of feeder sectionalization and reconfiguration problems deals with the integration of distributed generation (DG). The DG units provide a possibility for the sections to operate in islanded mode and protect the service for the customers of

the island. The prevailing purposes within this class are (1) optimal DG placement for a given recloser placement; (2) reconfiguration of feeders to support the efficiency of DG operation; and (3) increase of positive DG impact on grid operation.

The paper (Olamaei et al. 2007) addresses the DG efficiency and presents an approach where the objective function is cost summation of electrical energy generated by distributed generators (DGs) and substation buses (main buses). It is not equivalent to minimum power loss as the active power generation cost of DGs is considered as an optimal control parameter. Genetic algorithm is used to solve the optimal operation problem. The approach is tested on a real 20-kV distribution grid consisting of two feeders and two independent supply points, with two switches per each feeder. It was found that distribution feeders perform better with dispatched (DGs) rather than non-dispatched.

In some cases DG impact on distribution grid operation may be planned ignoring the recloser siting and reconfiguration issues. The paper (Mashhour et al. 2009) provides such an analysis by dealing with optimal siting and sizing of DG units for minimizing the total power losses in a radial distribution network under bidirectional and unidirectional power flow scenario. The DG may have a significant effect on losses. This effect can be detrimental or beneficial depending on the allocation of the DG units and their size. An Enhanced Genetic Algorithm is used to effectively explore the problem search space. Moreover, the article presents a simple and straightforward penalty function which does not require the normalization of the violations. The method is implemented and tested on a typical 16-bus distribution feeder. Simulation results indicate that unidirectional flow constraint may restrict the ability of DG units to minimizing the grid losses.

Reconfiguration is also often neglected when a grid planning is addressed, particularly, when the investment to conventional grid extension (without DG) is opposed to the investment for DG. Accordingly, the DG represents an alternative option to postpone the transmission and distribution expansion required to meet the forecasted peak loads. The typical study (Hussein et al. 2006) proposes an optimization model to determine optimal DG sizes and sites that minimize the new DG capital cost, DG operating costs, losses compensation costs, and cost of the purchased power by the distribution network. This optimization is performed subjected to several technical constraints related to both network and DG operation; it is also performed on hourly basis for different types of DG and for different scenarios. The investment required to upgrade the network without DG is compared with DG investments to justify their economical impact. The model was applied to a 31-bus feeder with several branches (laterals).

The paper (Chang et al. 2011) addresses the DG impact on grid performance (loss reduction) and DG penetration limits. The prerequisite of the analysis is the standpoint that the allowable DG interconnection capacity at each site of the existing network is often restricted by the current level of the fault, voltage variation and power flow constraints. An improved feeder reconfiguration technique is proposed to maximize DG penetration. The feeder reconfiguration problem involving the determination of DG positions, DG capacities and feeder tie switch locations is formulated as a mixed discrete nonlinear optimization problem that takes distrib-

ution security concerns into account. The solution of the problem is the identified locations of the tie switches. It was derived on the basis of crude load flow model and minimum voltage deviation criterion. The proposed particle swarm optimization (PSO) procedure was applied to improve the computational efficiency. A simplified grid consisting of three feeders and three ties (loops) was used to test the performance of the proposed method and compare it with the basic PSO-based method. The numerical results indicate that the proposed method can provide suitable feeder structure for loss reduction and accommodate higher DG integrations.

The study (Greatbanks et al. 2003) formulates and discusses a methodology for the optimal siting and sizing of DG under security constraints determined by voltage sensitivity and loss sensitivity analysis of power flow equations. The methodology takes into consideration a set of grid loading conditions and power factor levels. The authors note that, due to the potential limitations to the choice of sites, the optimal placement may not likely be applied in practice. The solution of optimal DGs sites is further used to optimize the system reliability applying the newly designed genetic algorithm. To be more precise, optimal recloser positions are sought out by minimizing the composite reliability index (CRI) as follows:

$$CRI = 0.2\frac{(SAIFI - 1)}{SAIFI} + 0.4\frac{SAIDI - 2.2}{2.2} + 0.4\frac{MAIFI - 7}{7}, \quad (1)$$

where *SAIFI*—system average interruption frequency index; *SAIDI*—system average interruption duration index; *MAIFI*—momentary average interruption frequency index. This methodology was validated by the numerical study based on 11-kV grid of one feeder with several laterals.

Original viewpoint on DG-enhanced distribution system planning under uncertainties was developed by Liu et al in Liu et al. (2011). The uncertainties refer to the stochastic output power of a wind generating unit, solar generating source and plug-in electric vehicle, volatile fuel prices used by DGs, and future uncertain load growth. To handle them, a new mathematical model was developed under the chance-constrained programming framework. The model yields optimal siting and sizing of DGs from the minimization of the DGs investment cost, operating cost, maintenance cost, network loss cost, as well as the capacity adequacy cost, respecting the security limitations as constraints. A Monte Carlo simulation-embedded genetic-algorithm-based approach was employed to solve the developed model. Finally, the IEEE 37-node test feeder was used to verify the feasibility and effectiveness of the developed model and method, and the test results demonstrated that the voltage profile and power-supply reliability for customers can be significantly improved and the network loss substantially reduced.

Like other recent studies, the study (Nerves and Roncesvalles 2009) gives preference to the evolutionary (genetic) computing techniques. It stems from the experience that many conventional optimization techniques, such as gradient methods, linear programming (LP), quadratic programming (QP) and dynamic programming (DP) may fail to find global optimum solutions when dealing with optimal location and sizing of embedded generation. The authors model the distribution system using a

load flow formulation wherein a distributed generator is modelled as injected real and reactive power. Evolutionary programming is implemented by perturbing distributed generator outputs and evaluating the fitness values (i.e. objective function values) of the resulting systems. Optimal locations and sizes of DGs are determined using a fitness value based on system loss reduction, while an optimal distributed generator schedule is determined for a 24-h period using a fitness value based on energy cost. Spot market prices are modelled on the basis of a typical 24-h price curve. The results for a 69-bus test grid of one feeder show that the method is effective and flexible.

The extended feeder was chosen as a test system to validate the ant colony system (ACS) algorithm in the problem of DG placement under fixed recloser locations (Wang and Singh 2008). Such a larger-sized distribution system was used to demonstrate the effectiveness of the proposed method that is known as having the ability to handle the discrete optimisation problem in various areas. The actual 394-bus and 1,123-node radial distribution network consisted of 199 loads, 104 laterals/sublaterals, and 44 normally closed enclosers. A composite reliability index was drawn up as an objective function to be minimized in the optimization procedure. This index C was derived through weighted aggregation of the following two indexes:

$$C = W_{SAIFI}\frac{SAIFI}{SAIFI_T} + W_{SAIDI}\frac{SAIDI}{SAIDI_T}, \tag{2}$$

where W_{SAIFI} and W_{SAIDI} indicate the weights for indexes SAIFI and SAIDI, while the subscript T indicates the target values. Furthermore, comparative studies with respect to genetic algorithms (GA) were also carried out. The feeder sectionalization scenarios with GA and ACS algorithm appeared to yield to different results.

2.4 Sectionalization and Reconfiguration of Grid with Allocated Distributed Generation

The last class of problems focuses on optimal recloser placement under given allocation of DG in distribution feeders, with eventual extension to simultaneous optimization of recloser and DGs placement.

The paper (Rugthaicharoencheep and Sirisumrannukul 2009) proposes a methodology based on fuzzy multiobjective and Tabu search to determine the optimal on/off patterns of tie and sectionalizing switches for feeder reconfiguration in a distribution system with DGs. Three main objectives considered in the feeder reconfiguration problem consist of power loss, feeder load balancing, and number of switching operations of the switches. The optimization problem is subject to power flow equations, voltage limits, feeder capability limits, radial configuration format, and no load-point interruption. All the objectives are fuzzified using a trapezoidal membership function. A max–min principle is applied to make a fuzzy decision for identifying a proper set of switching operation that offers a compromise among the three objectives. An illustrative case of a 69-bus distribution system, the same used in Rugthaicharoencheep

and Sirisumrannukul (2010) and Nerves and Roncesvalles (2009), demonstrates the performance of the methodology.

The paper (Li et al. 2008) suggests an optimization method to identify the optimum recloser placement in distribution grid to improve its reliability in situation with already allocated DG. To this end, the method may be used for planning new DG-enhanced feeder designs with an objective to maximize the reliability benefits achieved by the embedded DGs. This requires the use of a composite reliability index (2). Then the zone-network method is introduced for the reliability assessment. An improved genetic algorithm called Multiple-Population Genetic Algorithm (MPGA) is applied to search for optimum solutions. Using the MPGA, the optimization can be solved with multi-population: the influence of improper genetic parameters can be greatly decreased and premature convergence can be overcome effectively. Simulation was carried out for the 69-segment 8-lateral distribution feeder. For the future research, the authors aim to investigate the simultaneous placement of both enclosers and DGs, which are definitely interdependent.

Practically the same optimisation problem as in Li et al. (2008) is formulated in Wang and Singh (2006) and Wang and Singh (2008), i.e. how to find the optimal recloser placement for a DG-enhanced feeder. The same composite reliability index (2) is minimized for the 69-segment test feeder. The ant colony system algorithm performed effectively in both studies.

2.5 Problem to be Solved

Having reviewed the methodological advance and smart grid context of distribution feeder sectionalization and reconfiguration, we conclude that high combinatorial techniques are still dominating. This means that sectionalization and reconfiguration procedures need multiple scenarios of recloser placement (in sectionalization) and switch status selection (in reconfiguration), with the subsequent comparison of scenarios to identify the optimal one. The aim of this analysis is to propose a straightforward optimal sectionalization method for the specified distribution architecture.

3 Grid Architecture for Validation of Suggested Method

The suggested sectionalization method is applicable, at least, to the grid architecture presented in Fig. 1. Such an architecture consists of three radial overhead feeders, each fed from different independent supply point, usually a medium voltage bus of a sub-transmission substation. Herein buses are labelled A, B and C serving, accordingly, *feeder-1, feeder-2* and *feeder-3*. There are no lateral ties between the feeders.

The distinctive feature of the architecture in question is a connectivity of feeder end points. The points *ax, bx* and *cx* are located in close proximity and can be connected to 3-radius star by installation of short terminal ties (Fig. 3).

Each feeder is protected by a circuit breaker (e.g., CB1 on *feeder-1*) installed in its head part, just next to the substation bus. As Fig. 1 shows, a fault occurring at any location across a feeder triggers the opening of its circuit breaker and interruption of service for all of its loads. If a fault is not self-clearing, it is assumed to be eliminated by the repair crew affiliated to distribution network operator. If abnormal situation occurs, feeders cannot be coupled, i.e. switched over to another independent supply point.

4 Interruption-Cost-Based Sectionalization Method

The method provides a direct search procedure of optimal sectionalization. The optimality is referred to as minimum grid cost of non-distributed energy (NDE provided by the unique set of sectionalizing points on the feeders. This minimum corresponds to maximum saving of grid reference cost. Here NDE cost either before sectionalization (Fig. 1) or after sectionalization by other method (e.g. reason-based sectionalization) might be considered as reference cost. The direct search means that sectionalizing points are sought out by straightforward procedure, i.e. without choice from several options or iterative calculations.

The method deals with only permanent (long) interruptions, i.e. those which can be stopped only by the intervention of a repair crew. In this sense, the reclosers are considered as not performing autoreclosure function, but only isolating the faulted section. The method does not address the momentary and permanent interruptions (faults), which are triggered by transient disturbances and terminated by application of autoreclosure function.

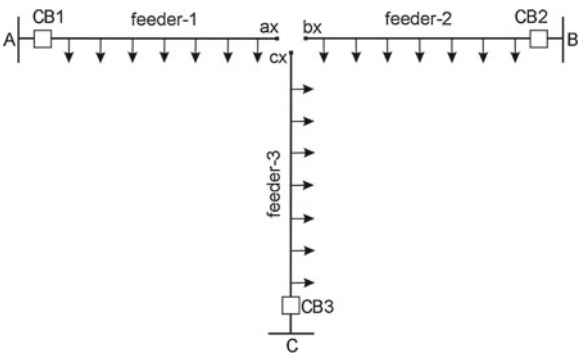

Fig. 1 Structure of method-applicable grid without switch-over possibility in the initial configuration

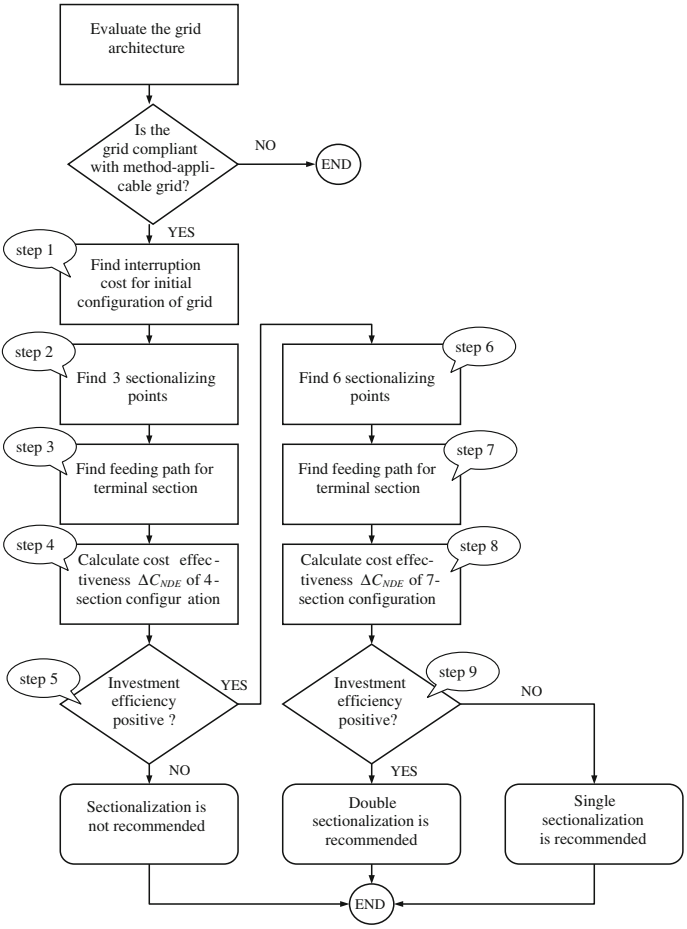

Fig. 2 Flow chart of suggested optimal sectionalization method for 3-feeder grid with three independent supply points

The Sects. 4.1–4.5 hereinafter specify the nine steps of the procedure to illustrate the suggested method. Figure 2 depicts the flow chart of this procedure, with references to steps 1–9. Procedure suggests undertake the single sectionalization (three enclosers, four sections) and then, if appropriate, proceed with the double sectionalization (six enclosers, seven sections)

4.1 Annual Interruption Cost Before Sectionalization (Step 1)

The method-applicable grid in Fig. 1 is given in the initial configuration, i.e. without sections. Thus, a fault, irrespective of its point of occurrence on a feeder, causes the opening of circuit breaker (e.g., CB1) and isolation of entire feeder (e.g., *feeder-1*, up to the end pint *ax*).

The annual interruption cost $C_{inter\ feeder}$ for each feeder is calculated as follows:

$$C_{inter\ feeder} = C_{NDE\ feeder} + C_{FC\ feeder}, \tag{3}$$

where C_{NDE}—grid operator's cost of non-distributed energy due to the interruption of distribution service in national currency units (NCU) per year; C_{FC}—fault clearing cost [NCU/year].

$$C_{NDE\ feeder} = \lambda_0 L t_{inter} T_{distr} \sum_{i=1}^{n} k_i \cdot P_i, \tag{4}$$

$$C_{FC\ feeder} = \lambda_0 L c_{crew}, \tag{5}$$

where

λ_0 : —failure rate per feeder [times/km/year]

L : —length of a feeder [km]

t_{inter} : —average single-fault-caused interruption time [min]

T_{distr} : —distribution tariff [NCU/kWh]

P_i : —load rating of *i-th* load [kW]

k_i : —load factor of *i-th* load

n : —number of loads on a feeder

c_{crew} : —average cost of repair crew ride to clear one fault [NCU/visit]

Finally, annual grid interruption cost C_{inter} is derived as a sum of feeder costs:

$$C_{inter} = C_{inter\ feeder-1} + C_{inter\ feeder-2} + C_{inter\ feeder-3} \tag{6}$$

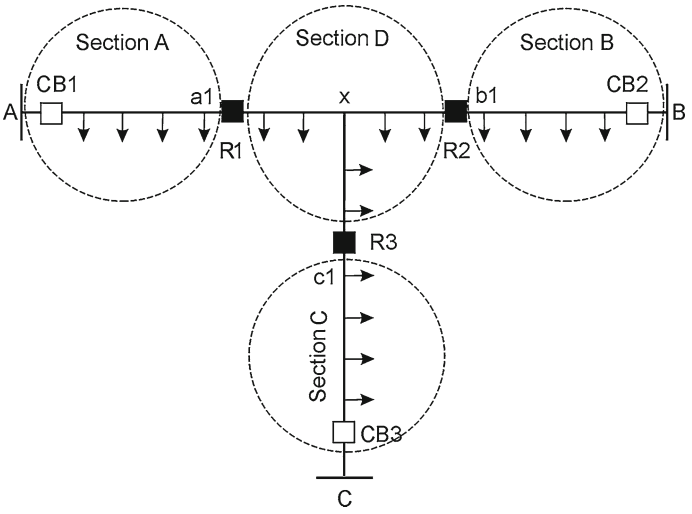

Fig. 3 Direct determination of three sectionalizing points in method-applicable grid

4.2 Single Sectionalization (Step 2)

The initial configuration is split to four sections A, B, C and D, each bearing the equal portion of grid interruption cost, i.e. 1/4 C_{inter} . The terminal section D is formed by linking the end points *ax*, *bx* and *cx* to a firm feeder intersection point x.

The sectionalizing points were found to be, say, $a1$ on *feeder-1,* $b1$ on *feeder-2* and $c1$ on *feeder-3*. The enclosers R1, R2 and R3 were placed on these points, respectively (Fig. 3).

4.3 Setting of Normal Configuration for 4-Section Grid (Step 3)

Contrary to the head sections A, B and C, the terminal 3-radius section D has no direct connection to the substation bus and can be fed only via a head section. The rationale for its choice is to ensure minimum joint SAIDI value for terminal and feeding head section. This means that the head section with least SAIDI is selected. The SAIDI of a head section is found as follows:

$$SAIDI_{section} = \frac{t_{inter} \sum_j N_{inter\ j}}{N_{section}}, \tag{7}$$

where $N_{inter\ j}$ is the number of interrupted customers in the head section under the fault j; $N_{section}$—the total number of customers connected to the head section.

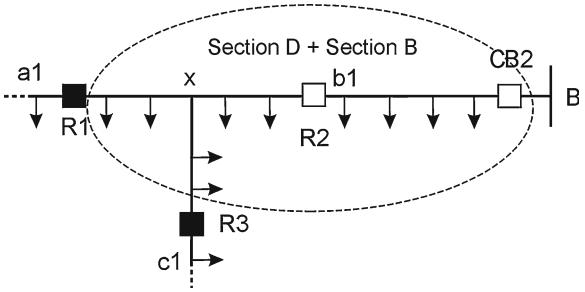

Fig. 4 Choice of feeding path to terminal section D in normal configuration

For instance, if section B has a least SAIDI value, it will provide a feeding path to section D in normal operation. To this end, the recloser R2 will be normally closed, while enclosers R1 and R3 will be open (Fig. 4).

4.4 Calculation of NDE Cost in 4-Section Grid (Step 4)

Using the same λ_0 value as in the initial configuration (see formula (4)), the annual NDE and its cost $C_{NDE}$are calculated in 4-section configuration appearing to be less because of the fact that sections A, B and C are shorter than feeders 1, 2 and 3, respectively. For instance, the failure of section A on *feeder-1* does not affect the operation of segment $a1 - x$ of the feeder,which would be impossible in the initial

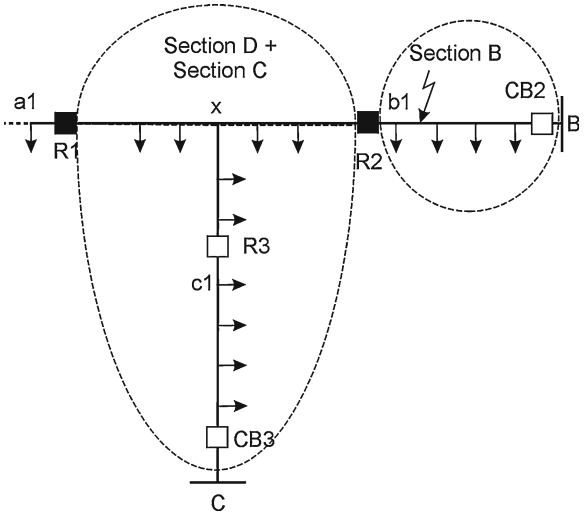

Fig. 5 Switch-over of the terminal section D in post-fault configuration

configuration. If a fault strikes section B (Fig. 5), the intact section D is switched over to section C immediately through the following automatic actions:

- R2 and CB2 open isolating the sections B and D,
- R3 closes providing a new feeding path to section D, the supply of which is restored via section C,

and, as a consequence, segment $b1 - x$ is further fed, differently from the initial configuration case.

The effectiveness of sectionalization can be estimated by comparing new cost value $C_{NDE\ opt}$ against the reference value $C_{NDE\ ref}$:

$$\Delta C_{NDE} = C_{NDE\ ref} - C_{NDE\ opt}, \tag{8}$$

where $C_{NDE\ ref}$ is either NDE cost value of the initial configuration (before sectionalization) or sectionalized configuration (established by other method than suggested one, i.e. its sectionalizing points differ from those determined in steps 2 and 3); $C_{NDE\ opt}$— optimal NDE cost value calculated according to steps 2 and 3; ΔC_{NDE}—effectiveness of sectionalization.

The effectiveness criterion ΔC_{NDE} is significant, yet not decisive. The feasibility of sectionalization is examined in step 5.

4.5 Efficiency of Investment for 3-Recloser Configuration (Step 5)

The economic performance of sectionalization shall be estimated in terms of investment efficiency, where investment (capital) cost is opposed to cost saving ΔC_{NDE} over the considered period. The investment-to-sectionalization cost INV_{sect} is expressed as follows:

$$INV_{sect} = INV_{purch} + INV_{inst}, \tag{9}$$

where INV_{purch} – recloser purchase cost, INV_{inst} – recloser installation cost. The efficiency indicators could be standard ones, such as net present value (NPV), internal rate of return (IRR) or/and payback period.

If the efficiency (NPV, IRR) is found positive, the sectionalization can be regarded as economically justified in the considered period.

The modern reclosers practically do not need maintenance and operation over the working life and do not incur the respective cost (Website of Tavrida Electric Website of Tavrida Electric).

4.6 Double Sectionalization (Step 6)

Further, having found that a single sectionalization is efficient, the opportunity to gain better efficiency through extended sectionalization could be examined. Its rationale assumes that smaller sections mean smaller isolated parts of grid and less numbers of cut-off customers.

Similarly to step 2, the initial grid (Fig. 1) is split to seven sections (also it could be said that the 4-section grid (Fig. 3) is resectionalized to seven sections using three additional enclosers), see Fig. 6.

- the sectionalizing points are found as those giving the sections equal portions of C_{inter} for the initial configuration (1/7 C_{inter});
- each feeder includes 2 sections and a tail participating in joint terminal section D:
  - sections A1 and A2 in *feeder-1*, with enclosers R1 and R2 placed at sectionalization points $a2$ and $a3$, respectively,
  - sections B1 and B2 in *feeder-2* with enclosers R3 and R4 placed at sectionalization points $b3$ and $b2$, respectively,
  - sections C1 and C2 in *feeder-3* with enclosers R5 and R6 placed at sectionalization points $c3$ and $c2$, respectively,
- feeder tails $a3 - x$, $b3 - x$ and $c3 - x$ are included into terminal section D.

4.7 Setting of Normal Configuration for 7-Section Grid (Step 7)

To set the configuration, a feeding path to terminal section D should be found. The search, as in 4-section case, targets at minimum SAIDI value of a path. Specifically, the enclosers R1, R4 and R6 are definitely closed.

Figure 6 illustrates the case when the feeding path $A - a2 - a3$ (*feeder-1*) is most reliable and therefore is linked with section D.

4.8 Calculation of NDE Cost in 7-Section Grid (Step 8)

By using (4), the annual NDE and new cost value $C_{NDE\ opt}$ are calculated. Certainly, they should be less than for 4-section grid—the fault striking, for instance, in a point just beyond the CB1 leads to isolation of section A1 (Fig. 6) which is shorter than the precedent section A in 4-section configuration (Fig. 3).

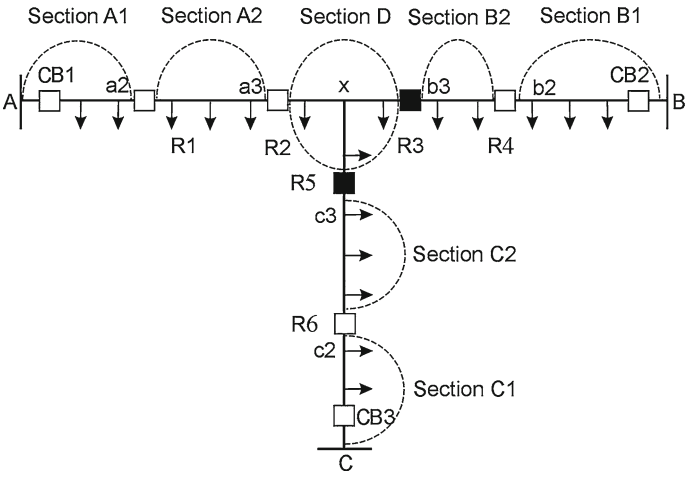

Fig. 6 Direct determination of six sectionalizing points in method-applicable grid

4.9 Efficiency of Investment for 6-Recloser Configuration (Step 9)

To be economically efficient, the increased saving ΔC_{NDE} should compensate the investment cost under double sectionalization. Here the same efficiency criteria as in step 5 shall be used to justify the doubled sectionalization.

5 Numerical Setup for the Validation of Suggested Method

The suggested method is further examined on the numerical setup based on 43-node medium voltage test grid presented in Fig. 7, with numerical data in Tables 1 and 2. The 10-year analysis period was considered, starting from the date of recloser installation. The periodical distribution tariff increment 0.006 NCU/kWh (5 % of initial tariff value) was assumed to take place each 2 years (Table 1). The grid is located in rural zone and services mainly small domestic customers. Numerical setup considers two geographic environments: forested and part-forested areas (Table 2).

All data are representative for Lithuanian critical load conditions as found in certain rural networks.

Specifically, the terminal point *dx* in the initial configuration (Fig. 7) graphically denotes 3 separate end points of feeders, as in the method-applicable grid in Fig. 7.

To appreciate the optimization advantage of the method, the effectiveness ΔC_{NDE} was derived from the comparison with the reason-based sectionalization. The notion "reason-based" is used as "somewhat heuristic", i.e. judged by expert experience.

Table 1 Test grid loading data in the numerical setup

Indicator	Value
Grid voltage (kV)	10
Total length L (km)	74.1
Rated load P (kW)	10000
Annual load factor of ith load k_i	0.33
Number of customers N	1165
T_{distr} (NCU/kWh)	0.12 – in 1st year
	0.126 – in 3rd year
	0.132 – in 5th year
	0.138 – in 7th year
	0.144 – in 9th year

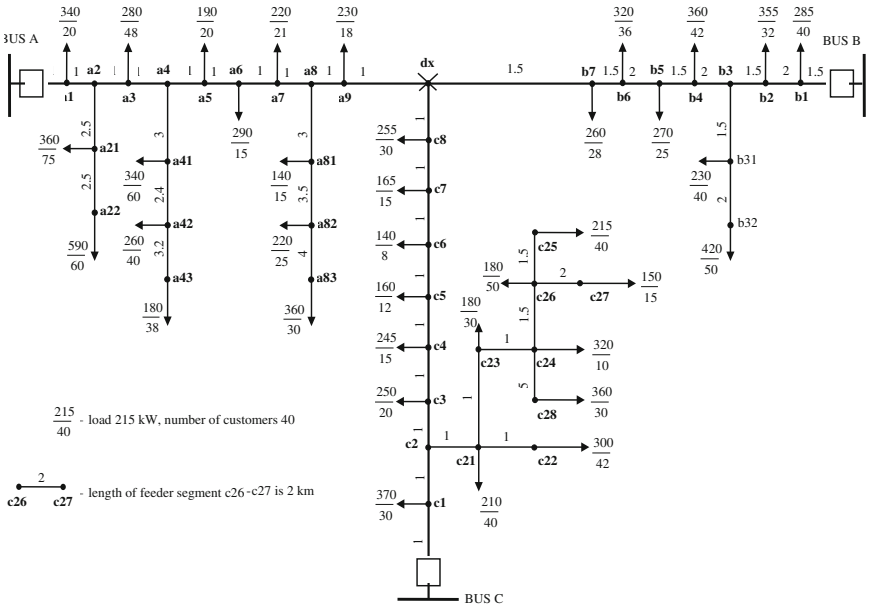

Fig. 7 Initial configuration (before sectionalization) of the test grid in the numerical setup

Table 2 Interruption data of the test grid

Indicator	Forested area	Part-forested area
t_{inter}(h)	3.4750	3.475
λ_0 (km^{-1}year^{-1})	1.59881	0.85
Number of faults (year^{-1})	118	63
c_{crew} (NCU/visit)	2000	1750

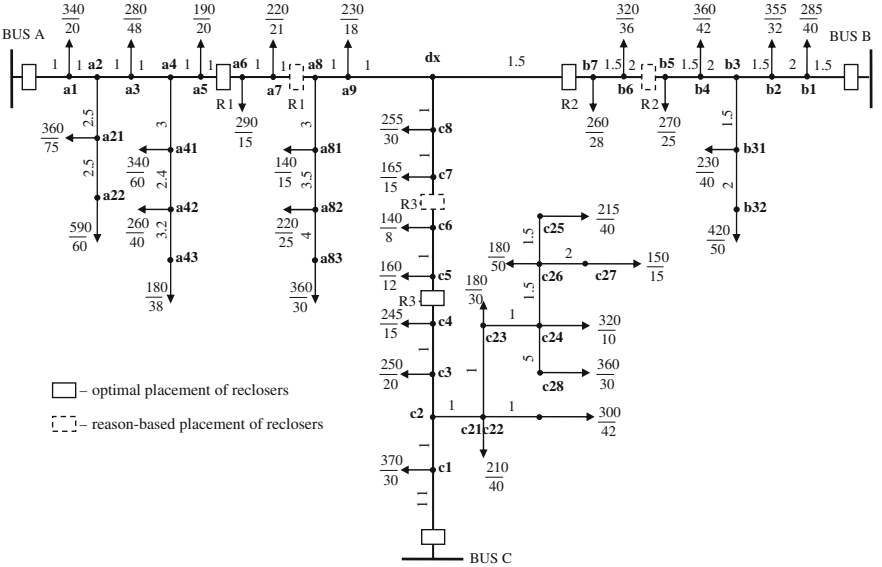

Fig. 8 Reason-based and method-based configurations of test grid under sectionalization to four sections

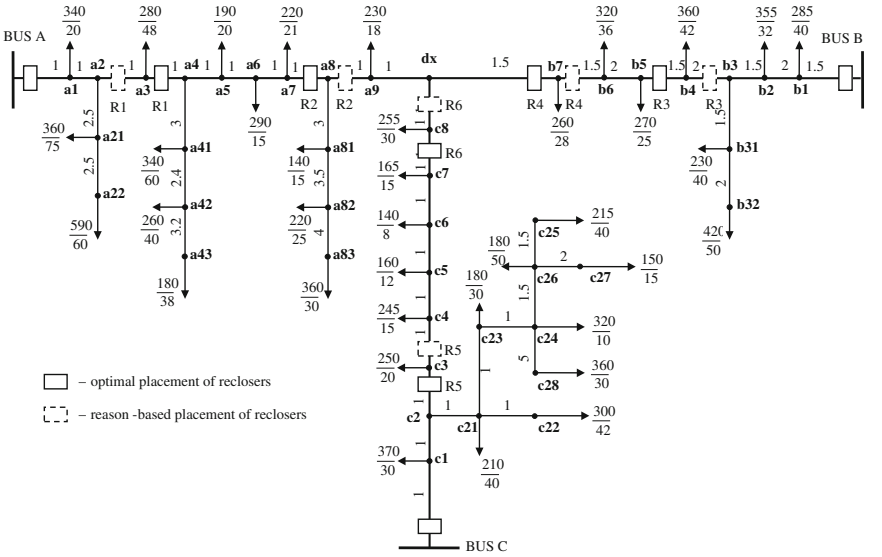

Fig. 9 Reason-based and method-based configurations of the test grid under sectionalization to seven sections

Table 3 Determination of equal portions of C_{inter} for the sections of test grid

First year cost	Forested area [NCU]	Part-forested area [NCU]	Comment
C_{NDE}	61423	30346	-
C_{FC}	236944	110224	-
C_{inter}	298367	140570	-
1/4 C_{inter} *ideal portion*	74592	35142	single sectionalization
1/7 C_{inter} *ideal portion*	42624	20081	double sectionalization

Table 4 Divergence of C_{inter} portions of the sections of test grid in forested area under single sectionalization

Section	Portion of C_{inter} [NCU/year]
A	75077 (*100.7 %*)
B	64919 (*87 %*)
C	74633 (*100.1 %*)
terminal	83738 (*112.3 %*)
ideal section	74592 (*100 %*)

The reason-based sectionalization was done prior to the method-based one. The sectionalization pictures are presented in Fig. 8 (4-section configurations) and Fig. 9 (7-section configurations) where the reason-based sectionalization is represented with enclosers in dash lines and method-based case—with enclosers in continuous-lines.

Specifically, the point *dx* in sectionalized configurations (Figs. 8, 9) graphically corresponds to the firm intersection point of three feeders in the terminal section. This section was linked to bus B via the feeding path $b1-b2$-…-$b7$, both in 4-section and 7-section configurations.

The ideal sectionalization would result in splitting the initial configuration to sections with equal portion of C_{inter}. It should be noted that practical sectionalization departs from ideal one because of the different lengths and load densities of the participating feeders. Since the test grid is highly asymmetrical in terms of feeder length (Fig. 7), real C_{inter} portions appeared to depart from ideal portions determined in Table 3. Such deviations are presented in Table 4 followed by the resulting variations of section loading data in Table 5.

Incidentally, the sectionalizing points (and recloser placement) under the method-based approach differed from those under the reason-based approach for each point (Figs. 8, 9).

The reason-based sectionalization appeared to be substantially less effective than the method-based one in terms of annual C_{NDE}. Tables 6 and 7 show their comparison where the cost values are discounted with the rate of 5 %.

As regards the entire considered period, the total NDE cost discounted with the rate of 5 % was compared with respect to the initial configuration in Figs. 10 and 11.

Table 8 presents accurate optimization effect over considered period.

Finally, the investment efficiency was calculated (Table 9) taking the following values:

$$INV_{purch} = 40,000\text{NCU/recloser},$$
$$INV_{inst} = 0.01 INV_{purch} = 400\text{NCU/recloser}.$$

The reliability before the sectionalization was very poor, and the placement of enclosers considerably mitigated such poverty (Table 10).

6 Discussion of Results

At first glance, the configuration of the 43-node test grid in numerical setup might seem strongly asymmetrical (Fig. 8) because total load of feeder B is taken by section B (recloser R2 is placed downstream the load *b*7), thus leaving no load for terminal section D. Having formally 3-radius structure, the latter section collects the loads

Table 5 Variation of section loading parameters in the test grid in forested area under single sectionalization

Section	Rated load P [kW]	Length L [km]	Load density [kW/km]	Number of customers
A	2540 (*25.4 %*)	18.6 (*25.1 %*)	136.6 (*99.1 %*)	361 (*30.9 %*)
B	2500 (*25 %*)	15.5 (*20.9 %*)	161.3 (*117 %*)	293 (*25.2 %*)
C	2780 (*27.8 %*)	18 (*24.3 %*)	154.4 (*112 %*)	322 (*27.6 %*)
terminal	2180 (*21.8 %*)	22 (*29.7 %*)	99.1 (*71.9 %*)	189 (*16.2 %*)
test grid	10000 (*100 %*)	74.1 (*100 %*)	137.85 (*100 %*)	1165 (*100 %*)

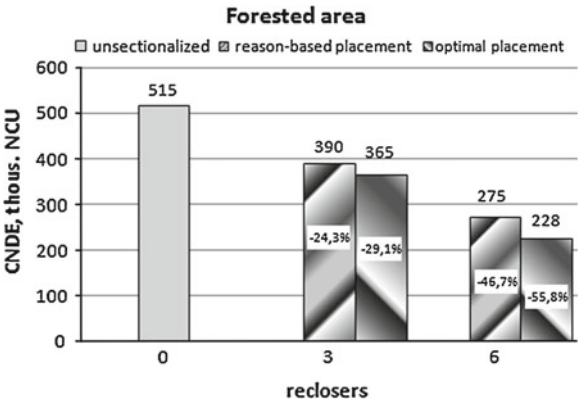

Fig. 10 NDE cost for the test grid in forested area over the 10-year period

Table 6 Reduction of annual NDE cost for the test grid in forested area when switching from reason-based to optimal sectionalization

Year	T_{distr} [NCU/kWh]	C_{NDE} in reason-based 4-section configuration [NCU]	C_{NDE} in meth-based 4-section configuration [NCU]	C_{NDE} in reason-based 7-section configuration [NCU]	C_{NDE} in meth-based 7-section configuration [NCU]
1	0.12	46481	43568	32755	27143
3	0.126	48805	45746	34392	28500
5	0.132	51023	47826	35956	29495
7	0.138	53242	49905	37519	31091
9	0.144	55460	51985	39082	32386

Table 7 Reduction of annual NDE cost for the test grid in part-forested area when switching from reason-based to optimal sectionalization

Year	T_{distr} [NCU/kWh]	C_{NDE} in reason-based 4-section configuration [NCU]	C_{NDE} in meth.-based 4-section configuration [NCU]	C_{NDE} in reason-based 7-section configuration [NCU]	C_{NDE} in meth.-based 7-section configuration [NCU]
1	0.12	22964	21525	16182	13410
3	0.126	24112	22601	16992	14080
5	0.132	25208	23628	17764	14720
7	0.138	26304	24656	18536	15360
9	0.144	27400	25683	19309	16000

Table 8 Optimization effect of the suggested method with respect to the reason-based sectionalization

Area character	Recloser placement	C_{NDE} [NCU/ 10 years]	Optimization effect [%]
forested	3 reclosers, reason-based	390462	−6.3
	3 reclosers, optimal	365993	
	6 reclosers, reason-based	275156	−17.1
	6 reclosers, optimal	228013	
part-forested	3 reclosers, reason-based	192907	−6.3
	3 reclosers, optimal	180819	
	6 reclosers, reason-based	135941	−17.1
	6 reclosers, optimal	112650	

only from 2 radii, i.e. $dx - a8$ and $dx - c5$. Nevertheless, if asymmetry is considered in respect of cost, it seems rather small. As Table 4 shows, sections A and C share ideal portion of grid interruption cost C_{inter} (practically, 100 % of value for ideal section), while sections B and D deviate from ideal portion by −13 % and +12.3 %, respectively. For better distinction, we suggest the density of the section load to be

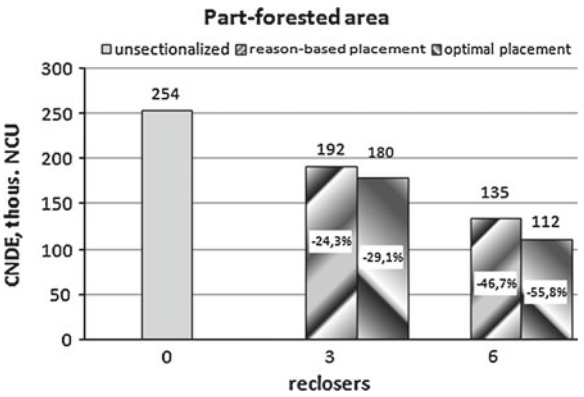

Fig. 11 NDE cost for the test grid in the part-forested area over the 10-year period

Table 9 Investment efficiency of sectionalization for the test grid

Area character	**Recloser placement**	**NPV [NCU/ 10 years]**	**Year of payback**
forested	3 reclosers, reason-based	4326	10
	3 reclosers, optimal	28794	8
	6 reclosers, reason-based	−1569	-
	6 reclosers, optimal	45574	9
part-forested	3 reclosers, reason-based	−5984	-
	3 reclosers, optimal	−47095	-
	6 reclosers, reason-based	−123418	-
	6 reclosers, optimal	−100127	-

Table 10 Reliability improvement in the test grid after sectionalization

Area	**Initial configuration**		**4 sections**				**7 sections**			
			reason-based		**method-based**		**reason-based**		**method-based**	
	SAIDI [min]	**SAIFI [times]**	**SAIDI [min]**	**SAIFI [times]**	**SAIDI [min]**	**SAIFI [times]**	**SAIDI [min]**	**SAIFI [times]**	**SAIDI [min]**	**SAIFI [times]**
forested	9368	42	6776	30	6580	29	5160	23	4189	19
part-forested	4628	22	3377	16	3251	16	2549	12	2070	10

an asymmetry measure as its variation is more pronounced: from 71.8 % (terminal section) to 117 % (section B) of the average of all sections (see Table 5).

The optimal sectionalization leads to significant reduction of NDE cost in the test grid. The single sectionalization gave as much as 30 % saving followed by 56 % under double sectionalization, both in forested and part-forested area (Figs. 10, 11).

The optimization effect against the reason-based sectionalization appeared to be considerable: 6.3 % of NDE cost saving under single sectionalization and as much as 17.1 % under double sectionalization.

Considering the investment efficiency (Table 9), the method-based sectionalization in forested area appeared to be efficient over the 10 years in both single and double sectionalization scenarios. Specifically, the reason-based sectionalization failed to reach a zero-efficiency in the latter scenario, although by a rather small margin. As for the part-forested area, the method-based sectionalization was absolutely inefficient in both scenarios.

As for reliability improvement, one can notice that method-based solution under single sectionalization yielded a small SAIDI improvement and inappreciable SAIFI improvement versus reason-based solution (Table 10). Under double sectionalization, reliability improvement was considerable, at least for SAIDI.

7 Conclusions

The distribution system sectionalization (recloser placement) and reconfiguration (closed/open status of enclosers) are receiving an increasing interest in the field of feeder design and operation. They are becoming an intrinsic feature of emerging smart grids by contributing to their reliability and self-healing capability.

Until recently, there is not a large number of sectionalization and reconfiguration methods proposed and the utilities often resort to the reason-based method which is based on practical knowledge of the specific distribution grid. Meanwhile the combinatorial or iterative methods require either more efforts for combination analysis or more time for running the iterations. Moreover, these methods have been validated for only certain (sample) grid architectures.

The suggested operator's interruption-cost-based sectionalization method is aimed at minimization of non-delivered energy (NDE) cost for 3-feeder distribution grid with terminal intersection node. It may be validated as an applicable, simple, and effective procedure, giving optimal sectionalizing point s for recloser placement on feeders. The method may be classified as direct search method. Contrary to other methods, it gives an optimal placement scenario at one straightforward calculation and gets along without the widespread comparison of many scenarios to identify the best one. It is not limited to a single or double sectionalization (with one and two enclosers per feeder, respectively) and might be extended to 3 and more enclosers per feeder.

The suggested method performed well when examined by a numerical setup based on Lithuanian data, with 3-feeder test grid covering 43 nodes in total. The minimum (i.e. optimal) NDE cost was found to be appreciably less than those provided by reason-based sectionalization. Saving of NDE cost in comparison with unsectionalized test grid was found to be sufficient to compensate the investment-to-sectionalization cost over 10 years period, both under single and double sectionalization. However, this finding is valid only for forested area. Nonetheless, there are some methodological opportunities for investment efficiency in part-forested area. One of them might be evaluation of saving in repair crew ride cost assuming that the sectionalization also brings down a number of permanent faults.

Another opportunity resides in saving of eventual penalties paid by grid operator to customers for service interruption. Furthermore, additional saving could be expected from increased smartness of sectionalization arising from the enhanced communications between enclosers (1) and extended reconfiguration actions, e.g. occasional supply of terminal section from two feeders simultaneously to carry on its bigger loads (2).

In broader context, the grid operator's interruption cost could be extended by customer interruption cost (i.e. customer losses). Then investment efficiency would be derived from combined social saving.

Additionally, the expansion of method's scope to other grid architectures, specifically those with lateral backup ties, can be planned for future investigations.

Acknowledgments This research was funded by a Grant (Nos. Ate—04/2012) from the Research Council of Lithuania.

References

Chang R-F, Chang Y-C, Lu C-N (2011) Feeder reconfiguration for accommodating distributed generations interconnection. In; Proceedings of 16th international conference on intelligent system application to power systems (ISAP), September 25–28, 2011, pp 1–6

Crow ML, Shetty N (2004) Electric power measurements and variables. In: Cleveland CJ (ed) Encyclopedia of energy, vol 2. Elsevier, Amsterdam, pp 245–254

Geynisman O, Schaub J (2007) Automated switching benefits customers. Transm Distrib World, September, 46–50

Greatbanks JA, Popovic DH, Begovic M, Pregelj A, Green TC (2003) On optimization for security and reliability of power systems with distributed generation. Paper Accepted for presentation at IEEE Bologna power tech conference, June 23–26, 2003, vol 1, Bologna, Italy, pp 1–8

Hashim AH, Mohamad AM, Abidin IZ, Baharuddin MZ, Yeoh EC (2006) Determination of autorecloser location using cost analysis in the Sabah electricity distribution network. In: Proceedings of 1st international power and energy conference, PECon 2006, Putrajaya, Malaysia, November 28–29, 2006, pp 587–590

Hussein DN, El-Sayed MAH, Attia HA (2006) Optimal sizing and siting of distributed generation. In: Proceedings of 11th international Middle East Power systems conference (MEPCON), El-Minia, Egypt, December 19–21, 2006, vol 2, pp 593–600

Jamali S, Shateri H (2005) Optimal application of reclosers and sectionalisers to reduce non-distributed energy in distribution networks. In: Proceedings of 18th international conference on electricity distribution, CIRED 2005 (June), pp 6–91–6. Turin, Italy

Jamali S, Shateri H (2005) Optimal siting of recloser and sectionalizers to reduce non-distributed energy. In: Proceedings of transmission and distribution conference and exhibition: Asia and Pacific, (2005) IEEE/PES, August 18, Dalian, China, pp 1–7

Liu Z, Wen F, Ledwich G (2011) Optimal siting and sizing of distributed generators in distribution systems considering uncertainties. IEEE Trans Power Deliv 26(4):2541–2551

Li Z, Yuqin X, Zengping W (2008) Research on optimization of recloser placement of DG-enhanced distribution networks. In; Proceedings of 3rd international conference on electric utility deregulation and restructuring and power technologies, DRPT 2008, Nanjing, China, April 6–9, 2008, pp 2592–2597

Mashhour M, Golkar MA, Moghaddas-Tafreshi SM (2009) Optimal sizing and siting of distributed generation in radial distribution network: comparison of unidirectional and bidirectional power

flow scenario. In: Proceedings of IEEE Bucharest power tech conference, Bucharest, Romania, Jun 28–Jul 2, 2009, pp 1–8

National Energy Technology Laboratory (2007) A systems view of the modern grid. Report (white paper) conducted for the U.S. Department of Energy Office of Electricity Delivery and Energy Reliability, January 2007

National Energy Technology Laboratory (2010) Anticipates and responds to system disturbances (self-heals). Report DOE/NETL-2010/1421, September 2, 2010

Nerves AC, Roncesvalles JCK (2009) Application of evolutionary programming to optimal siting and sizing and optimal scheduling of distributed generation. In: Proceedings of TENCON IEEE region 10 conference, January 23–26, Singapore, 2009, pp 1–6

Olamaei J, Niknam T, Gharehpetian G (2007) Impact of distributed generators on distribution feeder reconfiguration. In: Proceedings of IEEE Lausanne power tech conference, Lausanne, Switzerland, July 1–5, 2007, pp 1747–1751

Rugthaicharoencheep N, Sirisumrannukul S (2009) Optimal feeder reconfiguration with distributed generators in distribution system by Fuzzy multiobjective and Tabu search. In: Proceedings of international conference on sustainable power generation and supply, SUPERGEN '09, Nanjing, China, April 6–7, 2009, pp 1–7

Rugthaicharoencheep N, Sirisumrannukul S (2010) Feeder reconfiguration for loss reduction in three phase distribution system under unbalanced loading conditions. In: Proceedings of 45th international universities' power engineering conference, UPEC (2010), Cardiff, Wales, UK, Aug 31– Sept 3, 2010, pp 1–6

Wang L, Singh C (2008) Reliability-constrained optimum placement of reclosers and distributed generators in distribution networks using an ant colony system algorithm. IEEE Trans Syst Man Cybern Part C Appl Rev 38(6):757–764

Wang L, Singh C (2006) Reliability-constrained optimum recloser placement in distributed generation using ant colony system algorithm. In; Proceedings of IEEE power systems conference and exposition, PSCE '06, Oct 29–Nov 1, 2006, pp 1860–1865 Atlanta, GA, USA

Website of Tavrida Electric, group of Russian companies. http://www.tavrida.ru/Product/VacuumRecloser_CalculationStation/RecloserPbaTel/ (in Russian)

Wu W-C, Tsai M-S, Hsu F-Y (2007) A new binary coding particle swarm optimization for feeder reconfiguration. In: Proceedings of international conference on intelligent systems applications to power systems, ISAP 2007, November 5–8, 2007, pp 1–6

The Role of Flexible Demands in Smart Energy Systems

Kristin Dietrich, Jesus M. Latorre, Luis Olmos and Andres Ramos

Abstract The demand side of the electricity system holds a flexibility resource which has been ignored for a long time. With the presence of smart grids and facing challenges such as the massive integration of renewable energies into the system, demand side measures become viable and indispensable. This chapter will give a brief introduction about the concept of demand response. It will give an overview on demand response mechanisms, their objectives and potentials. Furthermore, an overview about various flexible demands in households, commerce and industries is given.

Keywords Demand side management · Demand response · Peak shaving · Demand shifting · Flexible demands · Power system modeling

1 Requirements of Smart Energy Systems

Smart Energy Systems must be flexible in order to adapt to quickly changing system conditions. The largest challenge on the way towards smart energy systems is the integration of renewable and distributed energies due to various reasons. On the one hand most renewable generation, and first and foremost those with a high future potential, such as wind or for certain regions as well solar generation, are intermittent. This implies variability in electricity produced by these generators and uncertainty in forecasts. On the other hand renewable units are far smaller in size and more decentralised than conventional thermal power plants. Additionally and on the contrary to thermal power plants, renewable power plants are less controllable. In absence of storage technologies natural resources as wind or solar energy can only be used when wind is blowing or sun is shining. Moreover, at the moment in many countries,

K. Dietrich (✉) · J. M. Latorre · L. Olmos · A. Ramos
C/Santa Cruz de Marcenado 26, 28015 Madrid, Spain
e-mail: kdietrich@upcomillas.es

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_4,

they savour priority in providing electricity when they are available. The rest of the generation mix has to adapt to it as they are little controllable. So, more flexibility is needed to integrate large quantities of renewable and distributed generation.

More flexibility can be achieved in a variety of ways. One way is to use more flexible thermal generation units. The conventional generation park may become more flexible by using more gas turbines, which can start-up quickly and are hardly constrained by their ramp rates. This potential is already being used and will go on to be exploited in the future. An upcoming problem is the high variable cost that gas turbines are subject to. Thus, they are situated at the end of the merit order. Very high or quickly changing intermittent energy production may create complications from a system operations point of view when the time of their occurrence coincides with low demand hours, i.e. when generation units at the beginning of the merit order are on-line. These generation units are base and middle load thermal plants, which, depending on the energy system, may be nuclear or coal fired power plants. They comprise generally the least flexible generation units. So, gas turbines are a great deal to manage peak demands but may be less useful in handling intermittent generation peaks in low demand periods. An alternative to introduce more flexibility in the system could be the use of storage facilities. They could store energy when electricity production by renewable sources is high and produce during demand peaks. But potent storage technology is either not viable (e.g. batteries cannot be used at large scale), not (yet) available (e.g. hydrogen storage is still not commercially available) or already exploited to a large extent (pumped storage hydro plants). Looking at the counterpart of the generation side whose potential has barely been exploited until now might be a solution to the missing flexibility problem: the consumption side has been considered until recently as completely inflexible. This has to be examined with more detail as some electrical demands are not at all inflexible and it is not a question of existing technology, but rather of economics and good regulation to use the existing potential.

In the following Sect. 2, we will give an insight into the objectives which Demand Side Management, and especially Demand Response, mechanisms pursue (Sect. 2.1), estimate the potential of certain DR objectives (Sect. 2.2) and classify and explain some Demand Response mechanisms in detail (Sect. 2.3). We will then continue with a more precise view on specific types of flexible demands apt to be applied for the formerly mentioned mechanisms in Sect. 3. International Experience is revised and an Outlook given in Sect. 4 and the following.

2 Demand Response Mechanisms

To explain in detail the potential of Demand Response (DR) and the mechanisms applied to achieve it, first the concept of Demand Side Management (DSM) has to be discussed. Demand Side Management comprises all activities which have the aim to change the demand profile in time or size. In contrast to DR, the concept of Demand Side Management has a wider scope. DR mechanisms are those activities

which are based on the reaction of demands to signals, especially price signals, while DSM includes as well information or education activities. We will focus on DR mechanisms as they include mechanisms which directly intend to change the demand profile and thus to provide higher flexibility to the system.

2.1 Objectives of Demand Response

While the objective of some DSM activities may be more general such as to sensibilise or to inform about the potential benefits of the reduction of energy consumption, DR mechanisms aim to change the demand profile directly in a way that causes less costs by responding to economical signals. That may imply to reduce demand peaks or to flatten the overall demand curve via valley filling or demand shifting. Or as well to make demands to be more flexible in general.

Peak shaving Shaving (or clipping) demand peaks implies reducing electricity consumption in high demand periods. This may be necessary in situations where demand exceeds available generation. This might happen in the case of a failure of one or more units during a demand peak. In this way peak shaving can avoid non-served energy. But in general, peak shaving may also be useful to flatten the demand curve in general to avoid serving peak demands, when costly generation plants are providing electricity.

Valley filling Valley filling means increasing consumption in off-peak periods. The objective is to handle more problematic situations in low demand periods. Electricity production by non-controllable generation could be higher than the instantaneous demand consumption. Valley filling could then relieve the overhang of generation without the need to spill valuable electricity. It is used as well to balance the daily demand curve during off-peak hours in contrast to peak shaving which acts during peak hours.

Demand shifting Demand shifting refers to the movement of demand from one time instant to another. Normally the shift will be from high to low demand periods. So, the daily load curve is flattened both during peak hours by demand reduction and off-peak hours by demand increase. But as well the contrary is possible in the case that renewable energies produce a lot of electricity during peak-hours and demand has to be increased. Depending on the demands, consumption can be delayed or advanced in time only some minutes, some hours or even between night and day.

Flexible load shape Another objective of DSM and especially DR activities is referred to as flexible load shape. This objective is related to reliability, and it refers to the ability of demands to react to sudden demand or generation changes in real time operation as reserve provider. The occurrence of a sudden failure in thermal units, might make the reduction of loads necessary. Increases in load might be required in case of lack of wind production due to forecast errors.

These and other DSM objectives are described in Gellings (1985) and Charles River Associates (2005) with more detail. The first mentioned objectives (peak shaving, valley filling and load shifting) are also known as load management objectives.

Table 1 Overview DR objectives

Load management objectives	Peak shaving, Valley filling, Demand shifting
Further objectives	Flexible load shape

Load is managed in such a way that complicated system situations are relieved. In Table 1 a brief overview is given. We will refer especially to peak shaving and demand shifting objectives in the following sections.

2.2 Demand Response Potential

The potential of demand to respond to system conditions has been fairly ignored until recently in Europe and until the end of the last century in the US electricity market. Flexibility came usually either from other, more flexible types of generation technology such as gas turbines, or from using some type of storage as for example pumped storage hydro plants.

So, the greatest part of the potential to use demand response is still to be exploited. In the literature this potential varies over a wide spectrum depending on the region, the type of consumption and the applied DR mechanism (see Sect. 2.3). The DR potentials mostly cited in the literature can be differentiated into two. The first describes the maximum possible reduction of demand during peak hours, the peak shaving potential. The second potential refers to the maximum amount of energy, which can be balanced through some hours by delaying or advancing certain loads, the demand shifting potential. A third potential commonly mentioned is the potential to decrease overall energy consumption by applying energy efficiency measures. But we will focus on the first two potentials as those are more important regarding the flexibility in the electricity system. The peak shaving potential ranges in the literature on average from 3 %(Faruqui and Sergici 2010) to almost 25 % (Borenstein 2005) while the statements of the potential for demand shifting oscillate between 5 % (Stamminger 2008) and 20 % (Figueiredo et al. 2005). More detail can be found in the Sects. 2.3 and 3.

The main driver of exploiting this latent flexibility resource is the benefit of employing it and thus the associated cost of implementation. One of the questions raised by many studies (e.g. Faruqui and George 2005; Conchado and Linares 2010) is exactly the profitability of applying Demand Response measures. Do costs outweigh benefits? This is the fundamental question as a wide acceptance among utilities and consumers is rather disputable if costs are not recovered. Results from three utilities in California (USA) analysed in the mentioned work of Faruqui and George (2005) come to very different conclusions about this issue. Authors in Conchado and Linares (2010) conclude that costs of implementing DR mechanisms exceed by far the benefits in the case of Spain. However, the authors admit that further possible

benefits should be evaluated in detail. Summarising the studies it seems that benefits are so far quite low or non-existent, but that benefits are most probable to get higher in the future. This future depends on the advances of DR technologies and consequently lower implementation costs, on higher opportunity costs of not applying DR (especially with the integration of renewable energies) and on using additional functionalities with the DR infrastructure.

Costs can be differentiated into investment and operation costs. Investment cost refer to enabling equipment that needs to be installed in the electric device, metering technology and possibly an energy management system grouping all individual devices for one point of consumption (common among residential consumers). Furthermore the term investment cost includes the whole communication and control infrastructure which serves to communicate with the electrical grid and thus the system operator. On the other hand operation cost include the fixed and variable costs of using DR technology. Fixed costs may refer to monthly internet rates (communication from domestic consumers might use the internet) or the cost of a control centre where price signal and their response are bundled to communicate with the system operator. Variable cost consider mainly the opportunity cost or the value of lost load in the case of peak shaving. In Paulus and Borggrefe (2009) the authors show that for industrial consumers investment as well as fixed costs are of minor importance as especially the metering and communication equipment might be already installed. In contrast the value of lost load can be very important if the demand cannot be recovered later which is the case of peak shaving. The cost distribution among residential consumers is completely the opposite: investment and fixed costs are major components of the total costs faced when applying DR mechanisms. Authors in Conchado and Linares (2010) estimated costs and benefits for residential customers. They show that 56 %of the total cost of implementation of a demand side management system is due to the installation of automatic control in existing appliances and smart plugs in homes. The control and communication infrastructure amounts to another 28 %, smart meters to 15 %and only 1 %of total cost corresponds to the operation of the DR technology. In the work of Haney et al. (2009) an extensive overview of international experiences about costs and benefits of smart metering is presented.

2.3 Types of Demand Response Mechanisms

Demand Response mechanisms are widely classified into two groups: price driven and incentive driven DR mechanisms (Albadi and El-Saadany 2007; Braithwait and Eakin 2002; Spees and Lave 2007).

Price driven mechanisms include those that are also known as dynamic pricing. Dynamic pricing implies that the price is not fixed but variable as it may be in Critical Peak Pricing, Time of Use Pricing or Real Time Pricing. More options are possible such as Extreme Day Critical Peak Pricing and Extreme Day Pricing.

Incentive driven DR mechanisms comprise direct load control and emergency demand programs, which are voluntary. Additionally, interruptible programs and

capacity markets, which are usually mandatory and demand bidding programs are included in this category (Moghaddam et al. 2011). Consumers offering ancillary services can be accounted as well to this category. The last two mentioned mechanisms rely on a market structure.

The authors in Spees and Lave (2007) describe as well other measures. They mention on the one hand subsidies on bank credits for DR technologies, which have not shown the wished effects until now. On the other hand they argue that bill discounts are seen as doubtful regarding the effects caused and audits only reach a very limited number of customers.

We will focus only on the most common of the price and incentive driven DR mechanisms in the following part (see Table 2).

Price driven DR mechanism Critical Peak Pricing (CPP) assigns higher prices to specifically announced periods. This may occur on few occasions a year. Various implementation options exist: the critical peak period might be fixed or variable and consumers might be advised one day ahead or on the same day. In the work of Faruqui and George (2005) CPP implies that prices on critical days are five times higher than the standard price and six times higher than off-peak prices. The authors test two different options: first the critical peak period is fixed and consumers are advised one day ahead, second the critical peak period is variable and notification is on the same day. Consumers in the second option could choose to get the necessary equipment installed free of charge. For the first option critical peak reduction in energy was about 13 %among residential consumers while the second option caused a 25 %higher load reduction in peak periods. This was basically due to the free installation of enabling technologies (mostly smart thermostats).

Time of Use (TOU) pricing refers to a price structure where prices depend on the time of the day. This may include different prices for peak and off-peak hours or even more periods of the day. In Faruqui and George (2005) experiments are run with a TOU price during peak periods of 70 %over the standard rate and 200 %over the off-peak rates. TOU pricing caused an average peak reduction of almost 9 %among commercial and small industrial consumers and about 6 %among residential consumers. In the work of Jessoe and Rapson (2011) TOU pricing is applied to commercial and industrial users and two to three periods each day are designated as high, medium or low pricing periods. The authors state as a main drawback the coarseness of TOU time periods which detained it from being really effective.

Table 2 Overview DR mechanisms

Price driven	Critical peak pricing, Time of use, Real time pricing
Incentive driven:	
Voluntary	Direct load control, emergency demand programs
Mandatory	Interruptible programs, capacity markets
Market structure	Demand bidding programs, Ancillary services

Real Time Pricing (RTP) implies that prices are adjusted close to real-time (hourly/half-hourly or even smaller periods). Thus, RTP transfers prices and thus system information to customers almost without time loss. Effects of an introduction of real-time electricity pricing at household level are analysed by Allcott (2009). This work captures the habitual response to variations in mean hourly prices, the response to deviations from the mean price structure and the cross-hour parameter for substitution to another hour. In the study of Boisvert et al. (2004) the response of commercial and industrial customers with respect to RTP is analysed. Great differences in price responsiveness among customer groups and peak periods are found. Furthermore peak load reduction were observed to depend on the faced price differences (see as well Sect. 3.2). The author of Sioshansi (2010) applies RTP tariffs to smooth out daytime load pattern in order to increase the use of wind power as wind generation is often curtailed because of system restrictions. Under real-time pricing more wind is utilized in the system and a higher percentage of demand is served by wind. Benefits of applying tariffs such as Real-time-pricing and Time-of-use pricing in a system with high wind generation are shown as well in Finn et al. (2009) for the domestic test case in Ireland.

More details and test studies on CPP, TOU and RTP can be found in Charles River Associates (2005), Faruqui and Sergici (2010), Moghaddam et al. (2011), Newsham and Bowker (2010).

Incentive driven DR mechanism Direct load control uses an option to reduce, interrupt or even increase power consumption of electrical devices in remote control. Direct load control may be a means, which is used in other mechanisms such as interruptible programs or ancillary services (Callaway 2010). Authors who analyse direct load control with an automated response find that specific loads are able to react to many short as well as to less frequent prolonged curtailments, (see Kirby et al. 2008; Burke et al. 2005; Eto 2009; Huang and Huang 2004).

Emergency demand programs and interruptible services offer their participants some financial incentive to curtail load immediately in the case of a system contingency. This incentive may be a discount, bill credit or as well a penalty for not responding to the curtailment signal. Normally the number of hours as well as the power that can be curtailed are limited. In a capacity market program, consumers also have to curtail load upon request and will be penalised if they do not respond. In contrast to the interruptible service, capacity market programs are not offered by the load serving entity. The curtailment in Emergency demand programs is voluntary while it is mandatory in the other two mentioned mechanisms. Different experiences can be found in Charles River Associates (2005), Tyagi and Black (2010) and Aalami et al. (2010).

In demand bidding programs consumers provide load reductions at a certain pre-specified price. This might be convenient for customers as they can normally rely on a fixed tariff but receive a pre-defined price for curtailments when prices in the wholesale market are high and load reductions are realized. Authors in Khajavi et al. (2011) and Nguyen et al. (2011) give an insight in different implementations.

Demands may offer positive as well as negative reserves by being disposed to suddenly decrease or increase consumption. Many small demands may offer ancillary

reserves in a more reliable way than a large generation unit might do (Kirby 2003). More details and studies on how demand may offer reserve can be found in Kirby et al. (2008), Hirst (2002), and Behrangrad et al. (2010).

Comparison of DR mechanisms In the work of Borenstein et al. (2002) the effectiveness of some DR approaches (RTP, TOU, CPP, demand-reduction programs, and interruptible programs) is compared. TOU pricing in comparison to RTP pricing reflects the real wholesale electricity price only to a minor extent. The two drawbacks of CPP for the system, namely the limitation of price and hours that can be curtailed, are at least for the consumer an advantage. Demand-reduction has the problem of finding a reliable baseline from which to pay the realised reduction. This leads to the adverse selection problem as first consumers will participate who have a consumption that is lower than the baseline. Customers whose consumption surmounts the baseline are not very likely to participate as they would pay more. If the baseline changes with the consumer behaviour, the consumer might be discouraged to reduce its consumption (moral hazard problem). Interruptible programs are mostly only a more coarse form of RTP or CPP. The authors conclude that demand reduction programs are worse alternatives than dynamic pricing such as RTP or CPP. On the other hand it has to be considered for the provision of reserves, that system operators do not have a guarantee that reserve is provided under RTP and CPP (response is probabilistic) while they have under interruptible or demand reduction programs. So, a sufficiently large number of independent customers is necessary to ensure a response induced by price changes.

3 Types of Flexible Demands

The following chapter will concentrate on the types of electrical devices which are apt to be used for DR mechanisms. We will consider first domestic consumers and then commercial and industrial ones. This distinction is important as households will act significantly different from commerce or industries due to the mere quantity of demand consumed and consequently the different financial incentives to be used to mobilize the existing DR potential.

Although households might be analysed as one type of consumer, demand levels, load curves and the penetration of certain electric devices depend not only on the region but as well on many other factors (e.g. ownership of air conditioning, number of bedrooms or annual income, see (Faruqui and George, 2005) for more information). Section 3.1 will give an overview on common domestic appliances which have a certain DR potential. For industrial consumers the possible types of demands not only depend on the region but on the particular industry and its underlying production processes. Hence, Sect. 3.2 aims to point out the behaviour and possible use for DR mechanisms for some selected industry appliances.

A possible way of managing many small demands in an effective way is the concept of Virtual Power Plants (VPP). These VPPs aggregate and manage loads to participate in markets for example. On a domestic level, these VPPs would com-

municate with the residential energy management systems. In the case of industrial loads communication could be directly with the manager of the VPP. For further reading on that specific topic, we recommend Braun (2007) or Dietrich (2011).

3.1 Domestic Demands

Domestic electricity consumption is making up 31 %of total electricity consumption in Europe (Eurostat 2009) (more information in Table 3). An important part of this consumption is due to thermal loads or those that are in some way controllable. We will focus on some of these appliances, study their penetration levels for the case of Europe and explain which DR potential can be used. We refer mainly to the studies undertaken by Stamminger (2008), Paulus and Borggrefe (2009) and ITA (2007). Electricity consumption for household devices is primarily taken from European Commission (2005) and Bertoldi and Atanasiu (2009).

In the considered studies the potential of participation of these household appliances has been analysed for various DR measures. These DR measures intend to achieve various DR objectives such as peak shaving and load shifting (see Sect. 2.1). Peak shaving and energy efficiency potential have been considered via the application of total interruption of the working cycle, limitations in power consumption, the limitation to more efficient programs or the use of other decentralised electricity sources such as solar, CHP plants or district heating. Load shifting potential has been studied by delaying the start of the consumption process, prolonging some parts of the consumption process to delay later more energy intensive phases or the use of energy storage capacities.

Home appliances may be classified depending on their penetration level. Refrigerators and washing machines are present in more than 95 %of the European households. Over 70 % of the domestic consumers own ovens and heat pumps. Freezers, dish washers, tumble dryers and electric water heaters have a medium penetration level of 52, 42, 34 and 23 %, respectively. To a lower extent air conditioning and electrical heating (both 8 %) is common. These data rely on the study by Stamminger (2008). Washing machines, dish washers, tumble dryers and heating pumps are non-thermal loads. In contrast to the other appliances, whose DR potential comes from controlling and reusing their thermal storage, their usage time can be altered com-

Table 3 Overview electricity consumption per sector based on Eurostat (2009)

Sector	TWh	% of total consumption
Domestic	839.111	30.7
Commercial	769.947	28.3
Industrial	980.994	36.1
Other	128.870	4.9
Total electricity consumption	2.718.922	

pletely. By shifting the consumption backwards or forwards as well as interrupting it completely, these appliances have great DR potential. All the mentioned home appliances amount to a significant part of the overall domestic electricity consumption, for the case of EU-27 of around 60 %(Bertoldi and Atanasiu 2009; Eurostat 2009).

Further appliances such as induction cooktop and ironing robots, which are less common are analysed in the report by ITA (2007). Appliances whose proper operation method makes them inapt to participate in DR mechanisms such as brown goods (e.g. TV, Audio), grey goods (e.g. PC, video-games), small household appliances and lighting are analysed in ITA (2007), as well as in Bertoldi and Atanasiu (2009).

Now, the more common household appliances will be described and their DR potential analysed in detail. An overview is given in Table 4.

Refrigerator The most distributed home appliance is the refrigerator with a penetration of 106 %(Stamminger 2008). Refrigerators use an insulated box as containment at whose back a cooling device is situated. A cooling compressor is used to compress evaporated refrigerant vapour. Afterwards the vapour condenses releasing the heat and expanding back to the evaporator. So, main electricity consumption is used for the compressor. This compressor is active only 20–35 %of the time connected to the electrical grid when no load is added. Refrigerators and freezers amount to 15 %of total household electricity consumption. Refrigerators can bear a load shedding potential for very short time periods. But before an interruption of service is

Table 4 Overview controllable domestic consumers

	% of household penetration	% of total dom. consumption	DR objectives
Refrigerator	106	15	Load shifting
Washing machine	95	6	Load shifting, Load shedding, Valley filling
Oven and stove	77	7	Load shedding
Heat pump	70	[a]	Load shedding
Freezer	52	[b]	Load shifting
Dish washer	42	3	Load shifting
Tumble dryer	34	2	Load shifting[c], Load shedding
Electric water heater	23	9	Load shifting, Load shedding, Valley filling
Air conditioning	8	2	Load shifting
Electric heating	8	19	Load shifting, Load shedding, Valley filling

[a] Incl. in electric heating
[b] Incl. in refigerators
[c] Load shifting to minor extent

possible, temperatures have to be cooled down to lower temperatures to tolerate the disruption in the service without food quality degradation. So, load is shifted forward in time. The compressor starting time may be delayed or interrupted taking into account the temperature to maintain food quality. General efficiency has improved over the last years as more devices with a high efficiency (energy label "A" and upwards) have been bought. This improvement is expected to go on in the future.

Washing machine The second most common household appliance in Europe is the washing machine. It is made up of a drum which is filled up to a certain level with water and rotates. The washing process includes then immersing the clothes in the water, heating up of the water to a certain predefined temperature to start the washing phase followed by several rinsing phases and a final high speed rotation to dry the clothes to a maximal extent. The electricity is mainly used for heating up the water at the beginning and for driving the motor for the drum. Washing machines cause around 6.4 %of residential electricity consumption in Europe. Already an 8 %of the washing machines feature some kind of control option such as the start time delay (Stamminger 2008). The electricity consumption can be lowered in general if the hot water is either stored in an preceding cycle, the water intake is hot or the water is heated with other sources such as local solar or CHP plants. A limitation to energy efficient ("ECO") programs, the interruption of the washing cycle or a limitation of the power consumption (lowering of temperature) might be other measures. Washing machines can apply load shifting by using a start time delay which waits until a signal from an energy management system or directly from the electrical grid is received to start the corresponding washing program. Certain programs may also be interrupted ride through.

Oven and Stove Stoves and ovens contain a heating element which transfers heat via thermal conduction or radiation, respectively. 7.5 %of total domestic electricity is consumed by electric ovens and stoves. Peak shaving potential lies in interrupting the cooking process for very short time periods but not directly after beginning in the heating-up phase.

Freezer Freezers work in the same way as refrigerators do. Normal temperatures range from −18 °C among normal conditions to −25 °C. The same load shifting and energy reduction measures as for the refrigerator apply (see page 10).

Dishwasher Dish washer consist of a tub, which is filled with water to a certain extent. Water is then heated up and through rotating arms water is sprinkled over dishes to start the cleaning process. Several rinsing phases and a final drying phase are included in the dish washing process. Electricity is mainly needed for heating up the water and for the motor of the rotating arms. Each household with a dish-washer consumes around 241 kWh per year making up 2.7 %of total domestic electricity consumption (taking into account the penetration mentioned at the beginning of this chapter on page 10. As in other non-thermal loads the starting time of the dish-washing process may be delayed, interrupted or reduced to lower temperatures. The electrical heating of the water might be replaced in the same way as proposed for washing machines (see page 10).

Tumble dryer After a washing cycle in the washing machine, laundry may be tumbled dry. Therefore hot air is blown into the drum of the tumble dryer through

the wet clothes. The wet air is either removed through venting to the environment or condensed by cooling down. Most electricity is used for heating up the air, rotating the drum and the fans. An annual consumption of 251 kWh per household and a penetration of 34.3 %cause around 2 %of total domestic electricity consumption (Stamminger 2008). Load shifting is possible but is regarded to be used to less extent as washing the clothes is the directly preceding process. Furthermore the heating process may be interrupted to shift load but without interrupting the drum rotation. Limitation to energy efficient programs or reduction of the drying temperature is possible but would prolong the tumbling process. Efficiency gains can be achieved by heating up water, which then heats up the air, with alternative sources, the connection of heat pumps or the use of gas-fired heaters.

Air conditioning Air conditioning is used to adapt the room temperature to comfort levels. The underlying cooling process of the most common devices is the same as that in refrigerators and freezers. Although per household consumption is quite high, due to the quite low penetration level in the residential sector and to the limited time of use during the year, air conditioning is making up only around 2 %of total domestic electricity consumption. See Sect. 3.2 for more information on the use in the commercial sector. The same DR measures as for refrigerators and freezers may be applied. Additionally, once passed a certain temperature threshold, cool outdoor air can be used to cool the room down.

Electric water heating Electric water heaters are used to warm up drinking water to be used in manifold household applications mostly in kitchens (e.g. washing dishes) or bathrooms (e.g. showering). They can provide heat instantaneously or by using a storage. Electric water heatings may work in a centralised way distributing the heat to various devices (see heat pumps) or decentralised. An electric current flows through an electric resistance, which produces heat. Electric consumption of heaters is making up around 9 %(Stamminger 2008; Paulus and Borggrefe 2009) of total household consumption. A complete interruption of the power supply in charging (less than half an hour) as well as maintenance mode (for various hours) is possible. Load reduction is as well possible through lowering the desired temperature. The load shifting potential consists in delaying the beginning of the heating phase.

Electric room heating Electric heatings are working in a similar manner as electric water heaters do. The heat can either be transmitted to the environment directly or using a storage unit. Storage units have a core which is able to store the heat (in contrast to electric water heaters the storage element is not water). When room temperature goes below a given threshold room air is absorbed and warmed up while passing by the storage core before being distributed to the environment again. Especially night storage heating are considered nowadays as inefficient. Some countries are trying to minimise the use of this type of heaters in the future (see Stamminger 2008, as well as Paulus and Borggrefe 2009) as it may be responsible for about a third of the electricity consumption of the here considered home appliances. Room heatings with storage bear a load shifting potential as the heat charging process could be interrupted or even increased depending on outside and room temperature. An interruption, except during the warming up phase, or a limitation of power is possible for around half an hour (ITA 2007).

Heat pumps Heat pumps are used to circulate the heat between different appliances (mostly between boiler and radiator). Its electricity consumption amounts to more than double the amount of that of tumble dryer and dish washer and are thus one of the largest consumers among the home appliances (Stamminger 2008). They can be turned off when outside temperatures are moderate until room temperature cool down to a given limit. In that way they are able to shed their load (Paulus and Borggrefe 2009).

3.2 Commercial and Industrial Demands

Commercial and industrial demands make up a total of 64 %of total electricity consumption in the EU-27 of which over half comes from industrial consumers (Eurostat 2009). This section will give some insight on specific examples of commercial and industrial appliances with DR potential.

Commercial Demands Largest electricity consumers in the commercial sector are office lighting, electric space and water heating, ventilation and commercial refrigeration with 21.6, 19.7, 12.6 and 8.7 %, respectively. Circulators, pumps, cooking appliances cause from 6.8 to 5.3 %of commercial electricity consumption. Computers and air-conditioning are responsible for a minor share of the consumption (Bertoldi and Atanasiu 2009).

In Faruqui and George (2005) small and medium commercial and industrial consumers and their response to different dynamic pricing tariffs (TOU, CPP and RTP) have been analysed. Central analysed devices with DR potential have been air conditioning and thermostats. On weekdays up to 9 %peak reduction could be achieved applying a CPP tariff.

In Instituto de Tecnología Eléctrica (2008) residential as well as small commercial consumers have been analysed and their consumption is classified into low, medium or high consumption and whether it has a rather flat consumption throughout the year or winter and/or summer peaks. This survey for the Spanish Mainland studied the behaviour of commercial consumers such as hotels as well as bars and restaurants whose main electrical devices with DR potential considered were air conditioning and room heating.

Industrial Demands Electricity consumption in industry process depends very much on the specific underlying process. The potential of either shifting, shedding or simply reducing the demand might be quite high. Highest DR potential include those processes which face both high electricity demands and high specific cost for electricity in the overall production process (Paulus and Borggrefe 2009). The authors of Paulus and Borggrefe (2009) found five processes, production of chloride, aluminium electrolysis, mechanical wood pulp production, steel production and cement milling, where the DR potential is significant and accounted for almost 20 %of overall industrial electricity consumption for the case of Germany. Those processes use either electrolysis, mechanical refining, melting or milling. Only part of these processes could be used for load shedding (chloralkali and aluminium electrolysis, cement mills and electric arc furnace) while another part could be used for load

shifting (chloralkali electrolysis and wood pulp production). The DR potential in three of the five analysed sectors is already exploited to a large extent.

Other authors do not go into detail on the specific underlying processes or appliances but give general recommendations based on the undertaken studies.

Authors in Eissa (2010) divide the small industrialized and commercial consumers into various consumption classes. Those with flat load consumption curve should intend to shift load as far as possible. Those with a modulated peak and off-peak use should try to shift demand from the peak to the valley. Up to which extent this demand shifting is possible depends on the underlying process and has to be analysed for each load separately.

Authors in Boisvert et al. (2004) study the reaction of over 50 commercial and industrial consumers to different dynamic pricing tariffs. They find huge differences in price responsiveness depending on the consumer group and peak periods. Most response was due to shifting demand from peak to off-peak periods rather than shedding it. As a consequence they recommend to use different tariff designs and expose consumers differently long periods to high prices depending on their load shedding capability. These authors found as well that the response of peak load reduction depends very much on the price difference from peak price to the normal rate. Peak load reduction would be proportionally higher when the price difference was higher. This leads to the recommendation of applying CPP for industrial consumers.

Electric Vehicles Electric Vehicles (EV) are a consumption which is still not taken into account as the penetration of EVs is only starting. EVs might be a domestic consumption if owned by private persons or commercial or industrial consumption if belonging to a vehicle fleet of companies. Different types of EVs including pure battery EVs, extended-range EVs and plug-in hybrid are considered in the report of Hassett et al. (2011). The authors of this study estimate that in 2030 between 15 %(sensible estimate of EV uptake) and 50 %(most aggressive EV uptake scenario) will be electric most of which (88 %) will be vehicles with four wheels and up to 8 seats apart from the driver's seat. The scenario which assumes a sensible estimate of EV uptake is considered to be the most likely one. Another interesting study can be found in Duvall et al. (2011). Although EV penetration are estimated to be fairly low at least in the upcoming years, their DR potential might be important in the future as they are—within a certain range—very flexible demands. Taking into account the assumed specific consumption, the distance travelled each day and the forecasts taken into account for this study (including five EU-countries), they assume that EV will be responsible for 102 TWh of additional yearly electricity consumption in 2030. This corresponds to a 1.8 %(sensible EV uptake) to 7.3 % (aggressive EV uptake) of total electricity consumption in these five countries assuming a 1 %annual demand increase starting from consumption levels in ENTSO-E (2012). Charging of EVs should be organised smartly taking into account the electric system conditions to not increase demand peaks even more (Hassett et al. 2011; Duvall et al. 2011). EVs could be used as a flexible load for valley filling during night. The use of EVs as flexible demand could even go further and be used as a storage device: the battery might be used as energy storage to shift load from peaks to valleys as described in Ramos et al. (2011).

4 International Experiences

In the following section DR mechanisms implementation and experiences of some selected countries are summarised. For further information we refer to an online database about the potential application and use of DSM in many countries, presented in IEA (2009).

United Kingdom The United Kingdom is leading in DR potential regarding the number of devices apt to support DR (Torriti et al. 2010). Not only smart meters but also old meters which are combined with a new unit or simply display units which are clipped on devices exist in the UK market. As in most European countries, DR progams for industrial and large commercial consumers have been in place for quite some time. This includes mainly interruptible programs which do not send economic signals but make use of some payment for the case of interruption. Apart, this type of customers may negotiate a TOU tariff with the supplier. There are various TOU rates available for residential customers. One example is the Economy 7 tariff which offers during night hours a far cheaper rate than during the day. Especially for residential consumers with electrical heaters with storage (see Sect. 3.1 on page 12) this is an interesting option. Loads are allowed to participate in markets for reserves (frequency and voltage, spinning reserves and others) in an aggregated form (Heffner et al. 2007).

France In the Nineteen-nineties a very effective CPP program was introduced in France, the Tempo tariff. It distinguishes three types of tariffs indicated by colors: white, blue and red. While the white tarriff is applied the majority of days of the year, the blue on around 12 %of the days, the red one is applied on very few days. The red tarriff is significantly more expensive than the normal rates. Consumers are informed of the type of tariff one day ahead (Torriti et al. 2010).

Italy With the highest rate of smart meter roll-out in Europe, Italy may count to the European countries with the highest potential of DR use. Large industrial consumers are subject to interruptible energy programs which, depending on the program, shed load in real-time or with a 15-min notice margin (Torriti et al. 2010). TOU-rates are applied since some years with the aim of shifting load from peak to off-peak hours and will be extended consecutively to all consumers with smart meters.

Spain Although wind is considered to be the main driver of DR implementation in Spain by the authors in Chardon et al. (2008), only few DR mechanisms are currently operating in Spain. These do not include domestic customers who simply pay a flat regulated tariff consisting of an energy charge and a peak demand charge. Commercial consumers with peak demands over 10 kW are subject to Time-of-use tariffs. The author in Bañares Hernández (2008) estimates that 20 %of the total demand is subject to tou tariffs. The System Operator is offering as well an interruptible load program. Currently 155 customers are taking part in this program with a total interruptible load of 2174 MW (Red Eléctrica de España 2011), which amounts to almost 5 %of the peak load.

Nordic Region The authors of Heffner et al. (2007) provide some insight into the Nordic Region and how demand is taking part in the markets of Denmark, Finland,

Norway and Sweden. Loads can participate in the regional energy as well as the power market. Furthermore demands can provide reserves either through bilateral contracts or bids.

United States Demand Side Management and all type of DR mechanisms have a far longer history in the USA than in Europe. An overview of the initial experiences with DSM programs, especially in the USA, can be found in Charles River Associates (2005).

Some of the US markets will be mentioned in the following paragraph. In PJM there are currently seven DSM programs running (Walawalkar et al. 2010). An emergency DR program whose participation is voluntary and an interruptible service program which is mandatory for up to 10 events each year, each lasting up to six hours. Furthermore there is a demand resource program which makes consumers eligible to receive capacity revenues. In a day ahead DR program consumers must submit bids in day ahead energy market and face a penalty for non-compliance. In the voluntary Real Time Demand Response Program customers need to notify PJM of the intention to curtail load one hours before load reduction is performed. Ancillary services can be provided by demands in a market. DR resources are required to bid into ancillary service markets, and to respond to any event similar to how a generator would. Additionally, an energy efficiency program is running, which allows consumers to receive a capacity credit for a certain time span. The minimum size for all programs is 100 kW. It has been seen that the participation in Emergency programs is far higher than in economic programs (Walawalkar et al. 2010).

In Heffner et al. (2007) the authors describe among others how demand is providing reserves in PJM and ERCOT markets. ERCOT, the ISO of the state of Texas, allows loads to provide various kinds of reserves. Furthermore, load is allowed to provide half of the energy needed during a contingency.

In NYISO five DR programs are currently in operation (Walawalkar et al. 2010). These include reliability based DR programs such as an emergency DR program, which is a voluntary program, an mandatory interruptible service and a Targeted Demand Response Program to respond to local reliability events within a particular zone to avoid the need to activate emergency events for the entire zone. Economic DR programs allow customers to offer curtailment bids in a day ahead market. Furthermore a Demand Side Ancillary Service Program was initiated in 2009 and allows demands to provide three ancillary services to NYISO markets. In NYISO some of these programs have minimum sizes: 100 kW for emergency and 1MW for economic and ancillary services, respectively.

Although energy efficiency has not been the topic in this chapter because other DR objectives mechanisms are more related to the issue of smart grids, we refer to some literature for further reading: for a description of currently applied energy efficiency measures in Europe, Asia and Latin America, the work of Boshell and Veloza (2008) may be of interest. An evaluation of a multi-country study about energy efficiency in the new member countries as well as the EU-25 can be found in IEE (2009).

5 Outlook and Conclusions

Demand Response forecasts show a great potential for future DR use: from 174 TWh implementing the adopted measures to 407 TWh if implementing as well additional measures for EU-27 (European Commission 2005), or more than 13 GW instantaneous power for ten UCTE countries (Torriti et al. 2010). Load management programs which intend to shave peaks, fill valleys or shift demands from peaks to valleys as well as the general higher flexibility of demand are of great interest when coming to the topic of smart grids. Smart grids are integrating not only an increasing amount of intermittent renewable generation but also a high number of other distributed energies. More intelligence and flexibility are needed in the electric system to face these challenges. DR mechanisms can help with both flattening the demand curve to avoid costly demand peaks and boosting the integration of renewables through more flexibility in the system.

References

Aalami H, Moghaddam MP, Yousefi G (2010) Demand response modeling considering interruptible/curtailable loads and capacity market programs. Appl Energy 87:243–250

Albadi MH, El-Saadany EF (2007) Demand response in electricity markets: an overview. In: 2007 IEEE power engineering society general meeting, pp 1–5, June 2007

Allcott H (2009) Real time pricing and electricity markets, MIT CEEPR Center for Energy and Environmental Policy Research, Working paper 09-015

Bañares Hernández S (2008) El equilibrio instantaneo generación-demanda: comportamiento de la demanda y su participación en la operación del sistema eléctrico (The instantaneous equilibrium of generation-demand: the behavior of demand and its participation in the operation of the e. In: Jornada Técnica sobre Gestión de la Demanda eléctrica (Technical Workshop on Demand Side Management), Sevilla, organized by Red Eléctrica de Espana, pp 25–32

Behrangrad M, Sugihara H, Funaki T (2010) Analyzing the system effects of optimal demand response utilization for reserve procurement and peak clipping. In: Power and energy society general meeting, 2010 IEEE, pp 1–7

Bertoldi P, Atanasiu B (2009) Electricity consumption and efficiency trends in European Union, status report. Technical Report, Joint Research Centre (JCR)

Boisvert R, Cappers P, Neenan B, Scott B (2004) Industrial and commercial customer response to real time electricity prices. Technical Report, Neenan Associates

Borenstein S (2005) The long-run efficiency of real-time electricity pricing. Energy J 26:93–116

Borenstein S, Jaske M, Rosenfeld A (2002) Dynamic pricing, advanced metering and demand response in electricity markets, UC Bukeley, Center for the Study of Energy Markets, Working Paper CSEM WP 105

Boshell F, Veloza OP (2008) Review of developed demand side management programs including different concepts and their results. In: Transmission and distribution conference and exposition: latin America, 2008 IEEE/PES, pp 1–7

Braithwait SD, Eakin K (2002) The role of demand response in electric power market design. Technical Report October, Laurits R. Christensen Associates, Inc.

Braun M (2007) Technological control capabilities of DER to provide future ancillary services. Int J Distrib Energy Resour 3(3):191–206

Burke RB, Henderson MI, Member S (2005) Operating reserve in New England. In: Power engineering society general meeting, 2005. IEEE, vol 2, pp 1570–1574
Callaway DS (2011) Can smaller loads be profitably engaged in power system services ? In: IEEE power and energy society 2011 general meeting, pp 1–3
Chardon A, Almén O, Lewis PE, Stromback J, Château B (2008) Demand response : a decisive breakthrough for Europe. Technical Report, Capgemini, in collaboration with VaasaETT and Enerdata
Charles River Associates (2005) Primer on demand-side management. Technical Report February, prepared for The World Bank
Conchado A, Linares P (2010) Gestión activa de la demanda eléctrica doméstica : beneficios y costes. In: V Congreso de la Asociación Española para la Economía Energética (AEEE), Vigo, Spain, 21–22 Enero, pp 1–19
Dietrich K (2011) Using demand response and flexible generation to reduce uncertainties in system operation. In: 11th young energy economists & engineers seminar—YEEES, Madrid, Spain, 24–25 Novembre, pp 1–27
Duvall M, Alexander M, Maitra A, Saucedo D, Jungers B, Halliwell J, Entriken R, Davis M, Davis K (2011) Transportation electrification—a technology overview. Technical Report, Electric Power Research Institute (EPRI) and PacifiCorp
Eissa MM (2010) Demand side management program evaluation based on industrial and commercial field data. In: Proceedings of the 14th international middle east power systems Conference (MEPCON10). Cairo University, Egypt, pp 15–19, Dec 2010
ENTSO-E (2010) Monthly consumption data of all countries for a specific year (Database: 19.01.2012). https://www.entsoe.eu/resources/data-portal/consumption/
Eto JH (2009) Demand response spinning reserve demonstration—phase 2 findings from the summer of 2008. Technical Report 500, Ernesto Orlando Lawrence Berkeley National Laboratory
European Commission (2005) Doing more with less. Green paper on energy efficiency. Technical Report 6, Directorate-General for Energy and, Transport
Eurostat (2009) Supply, transformation, consumption—electricity—annual data. http://epp.eurostat.ec.europa.eu/portal/page/portal/energy/data/database
Faruqui A, George S (2005) Quantifying customer response to dynamic pricing. Electr J 18(4):53–63
Faruqui A, Sergici S (2010) Household response to dynamic pricing of electricity—a survey of the empirical evidence. Technical Report February, The Brattle Group
Figueiredo V, Rodrigues D, Vale Z (2005) Simulating DSM impact in the new liberalized electricity market. In: 9th Spanish Portuguese congress on electrical engineering (9CHLIE), Polytechnic Institute of Porto, School of Engineering, pp 1–8
Finn P, Fitzpatrick C, Leahy M (2009) Increased penetration of wind generated electricity using real time pricing & demand side management. In: IEEE international symposium on Sustainable systems and technology, ISSST '09, pp 1–6
Gellings C (1985) The concept of demand-side management for electric utilities. Proc IEEE 73(10):1468–1470
Haney AB, Jamasb T, Pollitt M (2009) Smart metering and electricity demand : technology, economics and international experience, EPRG Working paper EPRG0903, Cambridge Working Paper in Economics 0905.
Hassett B, Bower E, Alexander M (2011) Evaluation of the impact that a progressive deployment of EV will provoke on electricity demand, steady state operation, market issues, generation schedules and on the volume of carbon emissions. Technical Report February, Electric Vehicle Penetration Scenarios in Germany, UK, Spain, Portugal and Greece, MERGE WP 3 Task 3.2 (I) Part I of Deliverable D3.2
Heffner G, Goldman C, Kirby B, Kintner-Meyer M (2007) Loads providing ancillary services: review of international experience, Oak Ridge National Laboratory, Working Paper ORNL/TM-2007/060
Hirst E (2002) Reliability benefits of price-responsive demand. IEEE Power Eng Rev 22:16–21

Huang K-Y, Huang Y-C (2004) Interruptible load management to provide instantaneous reserves for ancillary services. IEEE Trans Power Syst 19(3):1626–1634
IEA (2009) Demand side management program, http://ieadsm.org, date accessed:02.03.2009
IEE (2009) Evaluation of energy behavioural change programmes (BEHAVE). Intelligent Energy Europe and ADEME. www.odyssee-indicators.org, www.mure2.com, http://www.energy-behave.net/
Instituto de Tecnología Eléctrica (2008) Análisis de clientes, Entregable P.T. 1.2, Proyecto Gestión Activa de la Demanda. Proyecto CENIT GAD. Technical Report
ITA (2007) Informe entregable E1.1: Análisis de las cargas, Proyecto Gestión Activa de la Demanda. Proyecto CENIT GAD. Technical Report, Instituto Tecnológico de Aragón, García-Fuertes, Eduardo
Jessoe K, Rapson D (2011) Commercial and industrial demand response under mandatory time-of-use electricity pricing, UC center for Energy and Environmental Economics, Working Paper WP-023
Khajavi P, Abniki H, Arani A (2011) The role of incentive based demand response programs in smart grid. In: 2011 10th international conference on environment and electrical engineering (EEEIC), pp 1–4
Kirby BJ (2003) Spinning reserve from responsive loads, Oak Ridge National Laboratory, Working Paper ORNL/TM-2003/19
Kirby B, Kueck J, Laughner T, Morris K (2008) Spinning reserve from hotel load response. Electr J 21(10):59–66
Moghaddam MP, Abdollahi A, Rashidinejad M (2011) Flexible demand response programs modeling in competitive electricity markets. Appl Energy 88:3257–3269
Newsham GR, Bowker BG (2010) The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: a review. Energy Policy 38:3289–3296
Nguyen DT, Negnevitsky M, Groot MD (2011) Pool-based demand response exchange—concept and modeling. IEEE Trans Power Syst 26(3):1677–1685
Paulus M, Borggrefe F (2009) Economic potencial of demand side management in an industrialized country—the case of Germany. In: 10th IAEE European conference, energy, policies and technologies for sustainable economies, Vienna, 7–10 September, pp 1–32
Ramos A, Latorre JM, Báñez F, Hernández A, Morales-espa na G, Dietrich K, Olmos L (2011) Modeling the operation of electric vehicles in an operation planning model. In: 17th power systems computation conference—PSCC'11, Stockholm, Sweden, Aug 2011
Red Eléctrica de España (2011) Servicio de interrumpibilidad. http://www.ree.es/operacion/servicio_interrumpibilidad.asp
Sioshansi R (2010) Evaluating the impacts of real-time pricing on the cost and value of wind generation. IEEE Trans Power Syst 25:741–748
Spees K, Lave LB (2007) Demand response and electricity market efficiency. Electr J 20(3):17
Stamminger R (2008) Synergy potential of smart appliances. Technical Report, Rheinische Friedrich-Wilhelms-Universität Bonn
Torriti J, Hassan MG, Leach M (2010) Demand response experience in Europe: policies, programmes and implementation. Energy 35:1575–1583
Tyagi R, Black JW (2010) Emergency demand response for distribution system contingencies. In: IEEE PES T&D, pp 1–4
Walawalkar R, Fernands S, Thakur N, Chevva KR (2010) Evolution and current status of demand response (DR) in electricity markets: insights from PJM and NYISO. Energy 35:1553–1560

Smart Grid Tamper Detection Using Learned Event Patterns

William L. Sousan, Qiuming Zhu, Robin Gandhi and William Mahoney

Abstract The functionalities of a Smart Grid include the use of a network of power management devices to control the demand-based power distributions, dynamic pricing, reliable power quality monitoring, self-healing, and other customer services within the grid. However, by the very nature of these functions, there arises a great need for system security as the network connection offers would-be hostile attackers an ideal target. A protection mechanism that is adaptive and forward looking for the types of attacks that are unaccounted in terms of the latent system vulnerabilities is explored in this chapter. The method of learning normal operations and monitoring for abnormal operations has the potential for providing increased security and tamper detection within the grid. This chapter reviews present day state-of-the-art behavior monitoring and anomaly detection methods along with a presentation of a new method for learning event patterns and detecting anomalies for the purposes of tamper detection.

1 Introduction

With the proliferation of the Smart Grid Technology, the promise of a more reliable, secure, and cost-effective power grid continues to develop along with the maturity of the technology. Elements of the Smart Grid provide for the means of

W. L. Sousan (✉)
Technical Support Inc., Omaha, NE 68137, USA
e-mail: blsousan@gmail.com

Q. Zhu · R. Gandhi · W. Mahoney
College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA
e-mail: qzhu@unomaha.edu

R.Gandhi
e-mail: rgandhi@unomaha.edu

W. Mahoney
e-mail: wmahoney@unomaha.edu

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_5,

efficiently balancing the demand and production of power, quickly responding to increased loads, self-healing, adjusting residential power consumption by allowing power companies to turn off particular home devices, variable-rate billing based on power demands, remotely connecting and disconnecting meters, and a communications network to distribute the grid's operational data in a fast and timely manner (Fang et al. 2011).

A key component of the Smart Grid is its SCADA (Supervisory Control and Data Acquisition) architecture including the communication network that serves as the backbone for distributing information regarding the current state of the power transmission and supply. However, this network may also be viewed as a lucrative target for would-be attackers to invoke attacks such as Denial of Service (DOS), man in the middle attacks, corrupting network packets, and others. These attacks could bring down the power supply for cities, national key infra-structure systems, and others.

In addition to the network, one of the main devices within the Smart Grid is the Smart Meter that monitors the power usage within the grid and provides continuous real-time updates of power usage. The ability to remotely read power consumption also exhibits an opportunity for nefarious users to tamper with the power meters for purposes of electricity theft by adjusting the measuring and reporting logic. These issues have triggered such measures as the use of embedding specialized circuitry for tamper detection. For example, the work in Megalingam et al. (2011) describes the integration of circuitry into the case screws so that if the case is removed, power to the microprocessor is removed and this helps to inhibit the ability to subvert the Smart Meter's logic.

The commercial sector, military, utility companies, manufacturing plants and transportation networks all rely on SCADA systems to control geographically distributed cyber-physical assets. Not surprisingly, the Department of Homeland Security has grown increasingly concerned about securing SCADA networks. This concern is due to the fact that SCADA control systems are primarily owned by private companies, and over time these systems have been assembled together with a patchwork of different devices and communication protocols. There has often been little concern for security. These networks were originally intended to operate in complete isolation from public and corporate networks. Assumptions about such isolation are no longer valid with increased usage of low cost and widely available TCP/IP based devices and communication protocols. For example, nearly 1,700 of the 3,200 power utilities have some type of SCADA system in place, and roughly one quarter of these utilities have no separation between the corporate network and the system control network (Lemos 2005), in order to allow real-time collaborative business processes among bulk energy producers. The U.S. electric power industry and other critical infrastructure components are becoming lucrative targets for terrorist attacks; they are now strategic installations in times of war (Siobhan 2009). Even though many military bases have separate SCADA systems in place to provide local power, they also remain vulnerable because they use commercially produced, potentially flawed SCADA system hardware and software. For instance the Human Machine Interface (HMI) in a SCADA system might be implemented as a web server. Clearly, some

portions of the U. S. infrastructure are operating in a very dangerous mode. External entities that may be able to gain access to control centers could turn off power. In other SCADA installations a hacker could reroute trains, create chemical spills, or shut down factories. For example, hackers recently disrupted railway signals in the Pacific Northwest and caused 15 min delays of the rail schedules (Zetter 2012). Thus, a national security concern exists on two fronts: the capabilities of the military and public infrastructure safety.

This chapter is organized as follows. Section 2 briefly describes the concept and examples of system behavior monitoring approaches. It is followed by Sect. 3 that describes our work on a novel method of tamper detection implemented in a system known as SCADA-Hawk. Section 4 provides an evaluation of the SCADA-Hawk system. The chapter concludes with Sect. 5.

2 System Behavior Monitoring

Various methods in literature have described ways to monitor power distribution and smart grid technology in terms of the identification of normal and abnormal patterns of behavior. However, the components to be monitored may vary based on the behaviors to be analyzed. For example, some systems may focus on observing network connections for possible intrusion detection (Peterson 2004; Hansen 2008). Others may observe the traffic amongst the RTU's (Remote Terminal Unit) and supervisory stations such as the work in Cheung et al. (2007). Still others may monitor various system settings for non-normal system behavior by the non-standard configuration settings (Oman and Phillips 2007), and others may monitor the behavior of individual SCADA components such as RTU's (Yang et al. 2006; Cheung et al. 2007). As a result, there exist categories of behavior analysis, and the proceeding items are exemplars of the different categories.

An example of the category of message analysis, the work in (Cheung et al. 2007) describes a model-based ID (Intrusion Detection) system focusing on Modbus TCP networks. Their system is based on the assumption that SCADA networks are generally static in nature, and exhibit consistent traffic patterns that use few protocols and networks. The system's objective is to determine the normal behavior and watch for abnormal behavior that may indicate the presence of an attack. An advantage of using model-based detection is its potential for detecting unknown type of attacks as compared to a signature-based method. Although promising, the model-based approach may have limitations due to the vast differences of SCADA systems and equipment. In Cheung's work, several characteristics are monitored for behavioral patterns with the first being the Modbus protocol function codes. In addition to the codes, the communication patterns between the Modbus nodes and central computer may also be monitored. Finally, the system watches for changes in the services the nodes provides.

A SCADA configuration information monitoring system for detecting anomalies is described by the work of Oman and Phillips (2007). The research concentrated on

monitoring SCADA devices configuration settings and events for intrusion detection purposes. That also included the monitoring of login attempts as well as SCADA-specific commands such as loading new firmware and adjusting user privileges. A signature was generated of each RTU that is used for determining normal operations in Oman and Phillips work.

3 SCADA-Hawk System

Our research involves the development of technology used for anomaly detection within SCADA-based systems such as the ones found with the Smart Grid. Our system, known as SCADA-Hawk^{TM} (Sousan et al. 2011), essentially establishes a perimeter of minefields amongst critical components within the SCADA system for the purposes of detecting abnormal behavior that may indicate the presence of system tampering or intrusion. Figure 1 outlines an abstract view of the SCADA-Hawk system.

The idea behind the SCADA-Hawk technology is to establish Electronic Security Perimeters (ESP's) around critical components within the Smart Grid that monitor behaviors exhibited by various devices for the presence of anomalies and reports any anomalies to the central monitoring station.

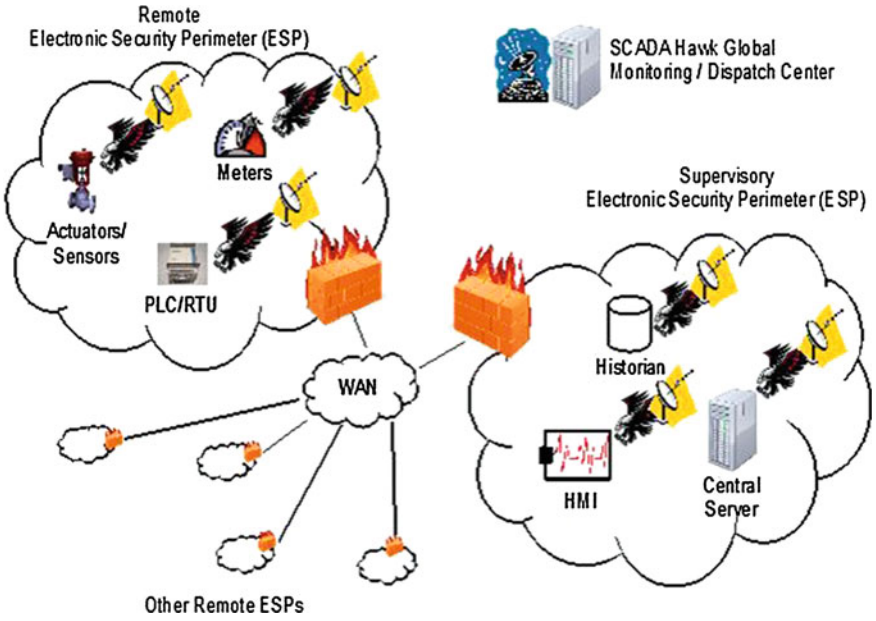

Fig. 1 Abstract view of SCADAHawk

The SCADA–Hawk functionality focused on the analysis of SCADA system's operational characteristics for developing a security system that is highly adaptable and flexible. Through a systematic learning process, we discover behavior patterns that repeat within the daily activities and functions of a given SCADA system. Our system abstracts the behaviors into events with corresponding time stamps and develops these timed events into models of event sequences using our "Snap-Shot" learning method and thus creates learned behavior models. Once the normal behavior patterns are learned and stored, the system can then be switched into a monitoring mode that compares current operations to previously learned models for anomaly detection. These anomalies may potentially indicate the presence of intrusions or attempts at tampering.

We have also developed a standardized means of modeling SCADA events by the development of a Taxonomy we call ComET (Common Event Taxonomy). This Taxonomy normalizes the multitude of signals generated for the various pieces of SCADA equipment into a Taxonomy that is a hierarchy of formally classified events along with their respective attributes.

For our current prototype, we developed a method of playing event sequences, both normal and abnormal, by the creation of an event player. This event player allows us to recreate the sequences of SCADA based operation events for evaluating SCADA-Hawk's learning, monitoring, and detection algorithms. The following sections describe the operations of the system.

3.1 System Design

The SCADA-Hawk system prototype is developed using Java language and NetBeans GUI builder along with several simulated hardware components of the system. The foundation of the system was to learn repeated SCADA events for the purposes of establishing relatively normal patterns of behavior and to continuously monitor the system for anomalies by always comparing the system operations to the learned models. To gain a better understanding of the security issues and types of events that would be suited for monitoring, we consulted several industry experts that use SCADA systems. Through several discussions amongst the experts and our team, we determined the types of events that we should focus and monitor. These are the non-functional (security-related) events such as login, firmware updates, file access, and I/O routines. These events appeared to be better suited for monitoring as opposed to attempting to measure the vast amounts of continuously fluctuating data points that occur as a result of operational and environmental changes.

3.2 COLLECTORs

The SCADA-Hawk system uses devices called Continuous Low Level Event Collection and Tamper-detection during Operational Realization (COLLECTOR)

for monitoring low level signals generated either from the hardware or software of a SCADA system. The COLLECTOR considers these signals as system-specific events or "Raw Events". These events can be generated from hardware by tapping onto various electrical points within the hardware that in turn will supply the COLLECTOR with signal from key parts of the circuitry and connections. For example, signals used for I/O control of sensors and actuators as well as signals used for configuration memory access may make ideal candidates for monitoring. In contrast to hardware points for monitoring, critical points in software may also be monitored. Software monitoring points are created by instrumenting the firmware to be monitored by attaching logic to report various executing methods and parameters for identifying the execution of critical operation code to the COLLECTOR. Thus the system to be monitored can be evaluated for critical operational points, i.e. potential target areas in both hardware and software, for COLLECTOR installation. The COLLECTOR's then continuously sense and report transitions in the system state.

Another key part of the COLLECTOR is the conversion of incoming signals from a "Raw Event" to their corresponding SCADA "Taxonomic" event by using the ComET Event Taxonomy (See Sect. 3.4). This process is used to "normalize" all processed events to standard events as described in ComET. These events are time-stamped and sent off to their corresponding AGENT (See Sect. 3.3).

Generally the system will consists of many COLLECTORS deployed throughout the system and each COLLECTOR is assigned to an AGENT in the system and each COLLECTOR will report all events detected to their assigned AGENT. A comprehensive system analysis performed by the domain experts installing the SCADA-Hawk system will identify the number of needed COLLECTORs and AGENTS and the corresponding configuration needed to secure the system to be protected.

3.3 AGENTs

A software component we call the Autonomous Gated External Non-intrusive Tool (AGENT) in SCADA-Hawk system is responsible for creating learned models of behavior as well as later comparing these models against future operational sequences. The AGENT runs in one of two modes, either learning or monitoring. Incoming ComET events received from the assigned COLLECTORs are captured and inserted into event/time based matrices that are either converted into learned models or compared against stored models. Learned model of behavior are stored permanently within each AGENT until a need arises to relearn behavior. While in monitoring mode, a sequence comparison is performed which verifies each incoming ComET pattern sequence against all the learned models. If a suitable match cannot be found with the current parameter configuration, and anomaly is reported to the monitoring process.

3.4 ComET

The Common Event Taxonomy (ComET) was developed to translate and normalize the various "Raw Events" monitored by the COLLECTORs into common SCADA events. This allows for a standardization of the SCADA events and their corresponding attributes from the myriad of equipment vendors into a common Taxonomy that provides an ontological base of the learning, monitoring, and detection functions. In addition, the integration of regulatory requirements such as those found in the North America Electric Reliability Corporation (NERC) into the Taxonomy adds an additional layer of events and attributes to be considered for security assessment of the SCADA systems.

The generation of events in a standard format is a prerequisite to developing robust pattern discovery algorithms that are device and domain independent. Correlating events (possibly different but related) at a system-wide scope across different devices is necessary to detect attack propagation paths. To this end, ComET describes security relevant system messages and state transitions in SCADA systems, and to facilitate the implementation of the tamper-detection modules (AGENTs). Essentially, the ComET will enable the extraction and reporting of event messages in a standard format across a variety of SCADA devices that are fitted with the AGENTs. The existence of a common event format will foster the development of repeatable and cross-domain solutions for SCADA cybersecurity.

The SCADA ComET is described using semantic web representation standards (OWL 2009). ComET is not a single hierarchical structure, but a multi-dimensional categorization of events in a SCADA system. This design decision allows us to capture the following dimensions of an event in the SCADA system:

[RR] Regulatory requirements relevant to the event (Abstract knowledge)
Relates an event to North American Electric Reliability Corporation (NERC) regulations and National Institute of Standards and technology (NIST) best practices.
[RS] Role of the event generating entity in the SCADA system (Mid-level knowledge)
Examples: Human Machine Interface (HMI), Historian, Control Server, Remote Terminal Unit (RTU), Programmable Logic Controller (PLC), Sensor, Actuator, etc.
[FR] Functional relation of the event (Low-level knowledge)
Examples: Login, Logout, Port opened, Port active, Service enabled, Service active, Coil open, Coil close, etc.
[DR] Device type of the event generating entity (Low-level knowledge)
Examples: Application, Operating system, Hardware or Firmware
[MR] Vendor/Manufacturer of the event generating entity (Low-level knowledge)
Examples: General Electric, VxWorks, CitecSCADA, Microsoft, Linux, VxWorks, etc.
[DO] Domain of the event generating entity in the SCADA system (Abstract knowledge)

Example: Electric, Wastewater, Water, Refinery, Pipeline, Telecomm, Environment control, etc.
[CP] Communication protocol relevant to the event (Low-level knowledge)
Example: Modbus, DNP3, Inter-Control Center Communications Protocol (ICCP), TCP/IP, etc.
[VU] Common vulnerabilities relevant to the event (Low-level knowledge)
Examples: Password sent in clear text, wireless connection not encrypted using strong algorithms, Shared user accounts, etc.

The above dimensions of the SCADA ComET also span multiple levels of abstraction (abstract, mid-level and low-level knowledge). This distinction allows an application or human to consume the events at a level of abstraction that is most appropriate for the task at hand. Each concept in a dimension is assigned a unique ID for cross referencing.

3.5 Snap-Shot Learning

The SCADAHawk system makes use of a novel learning process that is based on storing frames of timed event sequences called "snapshots." During the real-time operation of the system, the captured snapshots are transformed and normalized into event sequence models and stored for later use for anomaly detection. The learning and monitoring processes take place within the AGENT using the identified events reported by the assigned COLLECTOR's. The structure of the event sequence models are such that the COLLECTORS are represented as columns and the rows are partial event sequences within a two-dimensional matrix. Thus each element within the matrix is an event whose column identifies which COLLECTOR reported the event.

The event model learning process takes place by assigning each incoming event to the next available row for the COLLECTOR that reported the given event in a matrix known as the State Transition Event Buffer (STEB). Chronological order is represented in increasing time progressing from the top of the STEB to the bottom. As designated times the STEB is converted into Transition Definition Models (TDM's) by normalizing all the events times with respect to the initial recorded event. As a result, various event sequences are learned and stored as TDMs. Figure 2 shows the information flow for the learning process:

Note that the monitoring process is similar to the process above in that sensed Taxonomic events are captured and stored within a matrix known as the Current State Vector (CSV). In the monitoring mode, while each column represents the corresponding COLLECTOR that generated the event, there is only a single row to represent the current state of the system currently being monitored. Thus as each new event is received, it replaces the previous event contained within that element. This is because the CSV is, in essence, a partial sequence of the most recently received events from each COLLECTOR. The CSV as a sliding window moves along each stored TDM for the purposes of tracking a given event sequence. If a currently tracked event se-

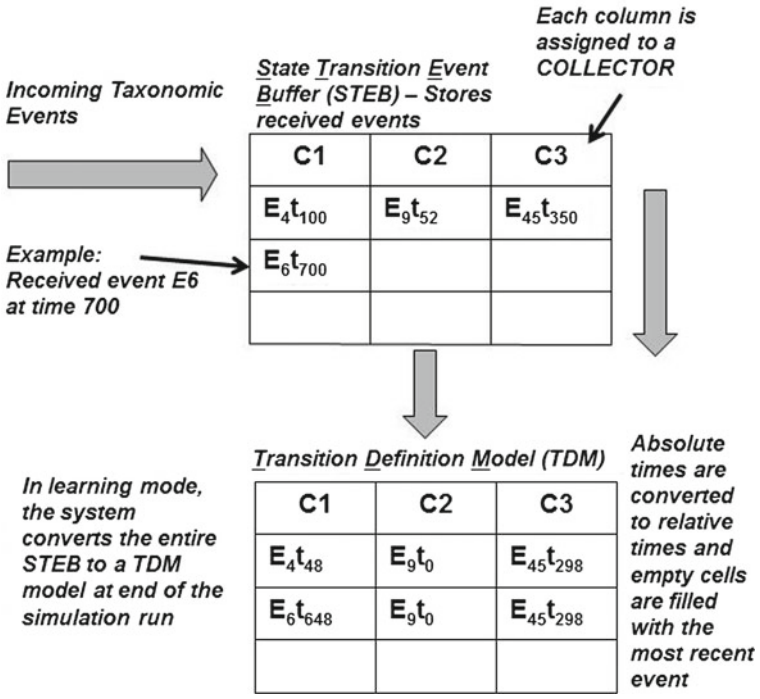

Fig. 2 Snapshot learning process

quence whose partial sequence is stored in the CSV cannot be matched to any of the stored TDM's, an anomaly is declared by the monitoring AGENT which then sends an alert. The basic outline of the monitoring logic is depicted in Fig. 3.

3.6 Event Player

A means for simulating SCADA-Hawk system operations was developed by the creation of an "Event Player" technology. This player allows users to describe the operational sequence of a SCADA system by defining a sequence of "Raw Events" and their corresponding times to generate each Raw Events. Both normal sequences and real-work attack vectors can be created and played for evaluating the algorithm's performance. These sequences can be defined by users within text files and then stored and later played back as needed. Furthermore, the sequence replay time could be increased or decreased by adjusting the virtual time representation.

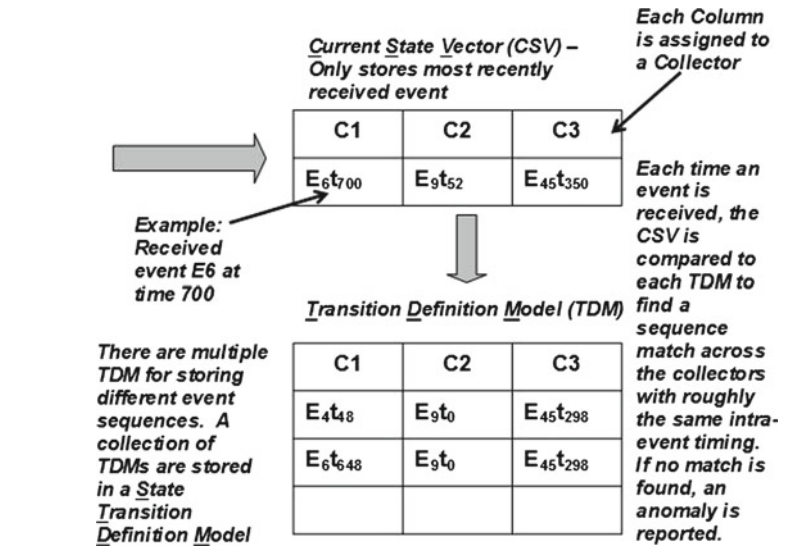

Fig. 3 Monitoring logic

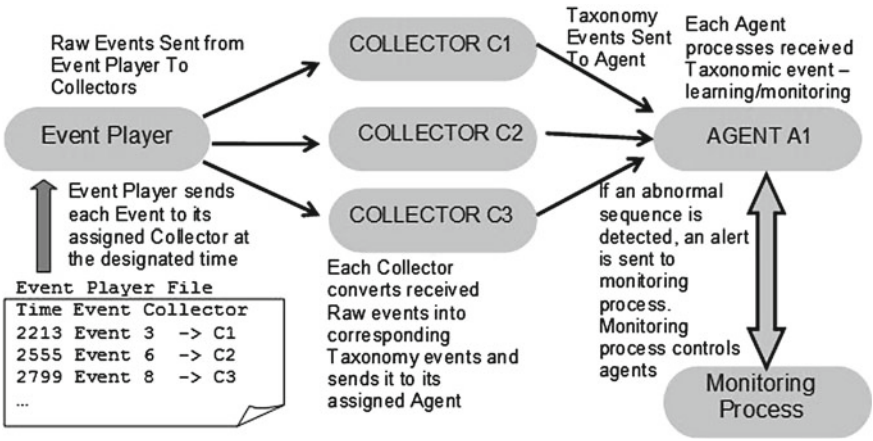

Fig. 4 Prototype operation

3.7 Prototype Operation

The developed technologies are all integrated into a Java application. The COLLECTORs and AGENTs were simulated as independent Java threads and the system contained provisions for configuring various combinations of COLLECTORs and AGENTs. An operational flow of the SCADA-Hawk prototype is depicted in Fig. 4 below.

The following sections describe the prototype operation.

3.8 Simulation Configuration

The SCADA-Hawk simulation system allows users to configure the number of COLLECTORs and AGENTS and their inter-connections and parameters based on the specific SCADA operations identified to be monitored. Based on a security and threat evaluation of the components and events to be monitored, the configuration can be tailored for monitoring identified raw events through a combination of observing both low-level hardware signals and identified software points through program instrumentation.

3.9 Simulate SCADA Domain Operation

After the COLLECTOR and AGENT configurations are complete for the particular SCADA domain to be monitored, the operations can be emulated through the use of the event player. Specific operational behavior can be emulated by creating corresponding event player files through detailing the event sequences within event player files. Sequences can be described by listing each event to be simulated along with its corresponding occurrence time and the respective COLLECTOR that has been identified to receive the event.

3.10 Simulation Execution

Once the monitoring devices have been designed and configured, and corresponding event player files developed, the system can be executed in one of two operation modes as mentioned previously. The first mode, learning mode, is used to play the normal behavior event player files to allow the system to learn the normal operation sequences of the SCADA domain under evaluation. While in learning mode, the AGENTs process the received events and build behavior patterns in preparation for later use in monitoring for anomaly detection. The second mode of operation, monitoring mode, uses the learned patterns of behavior to detect anomalies. This operational mode is used for evaluating the anomaly detection capability by replaying the event player files that contain potential attack vector sequences. In addition, variations of normal behavior patterns can be replayed to evaluate the system tolerance as to deviances amongst events in normal behavior. An alert is generated if a given monitored sequence does not match any learned sequences based on the current operating parameters. Either in learning or monitoring mode, various diagnostic files are generated that provide insight to the system's operation for use in evaluating performance and in refining the operating parameters.

4 Evaluation

Note that our system supports the ability to monitor low-level signals generated from hardware as well as monitoring strategic points with the execution flow of device firmware through program instrumentation. Thus our system can perform both "black box" and "white box" monitoring of a given system based on identified points to be monitored. Black-box Instrumentation strategy is used when limited privileges are available to modify the SCADA component. In this configuration SCADA-Hawk's COLLECTORs are attached to the component of interest at its points of entry and exit. This includes the following cases:

- (JTAG) IEEE 1149.1 Standard Test Access Port and Boundary-Scan Architecture
- Network interfaces (e.g. Packet sniffers):
- Alternation of signal on input and output wires (e.g. relay actuation):
- Peripherals (e.g. USB, serial ports, parallel ports etc.):

White-box Instrumentation strategy is used when privileges are available to modify the software process running on the SCADA component with an instrumented version. In this configuration the COLLECTORs monitor the basic blocks that are accessed during the execution of the SCADA process. This includes the following cases:

- Instrumentation of file access and network access processes in the OS.
- Instrumentation of the primary SCADA application.
- Memory bounds monitoring.
- System and application files integrity monitoring.

Initially, the types of threat models that exist and events (system mappings) need to be identified so as to determine how to configure the COLLECTORs and AGENTs for monitoring specific SCADA devices (i.e. RTU). As a result, the requirements for learning and monitoring are driven by the types of threat models and corresponding monitor-able events. The following model is used as a formalized process of determining attack vectors and COLLECTOR/AGENT configurations (Fig. 5):

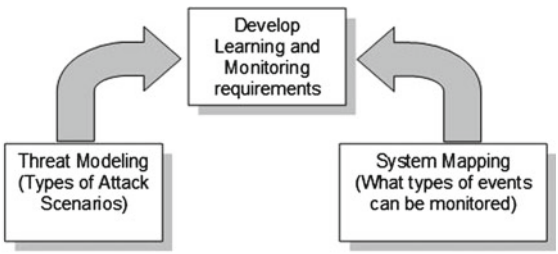

Fig. 5 Process of determining learning and monitoring requirements

We evaluated the features of the SCADA-Hawk system based on the potential attack vectors recognized with the vulnerabilities on an AMR (Automated Meter Reading) device. The device we experimented was a passive device in the field testing phase. It made a good candidate for the typical operations found in AMR devices. The tested operations included logic to track and record residential power usage over designated time increments, reporting of these records at pre-determined times, support variable rate billing based on time of power usage. The current model transmitted power usage records using the public phone network and enhancements are in progress for adding support for wireless communications.

The meter's hardware and operating functionality were analyzed to determine the types of threats that would be possible as wells as the types of motivation for potential attackers. From these analyses, two categories of threats were determined which are cyber-intrusion through the wireless communications and physical tampering by tampering with the actual meter. The following types of attack motivations where identified:

- Vandal/Insider—Financial Gain—alter the power usage reporting to show less usage, alter the power usage reading components so they report less power, alter power usage to show more usage in off-peak hours.
- Mercenaries—Vandalism—learn a resident's habits (i.e. when they are away from home opportunities exist for a break-in) by observing their power usage.
- Vandals—Nefarious purposes—cause meters to dial wrong number, cause meters to all dial out at once tying up phone lines/wireless connections and thus causing a Denial of Service (DOS) attack, alter power usage reading components to report false usage readings and thus confuse/stress the power grid supply needs.

We developed a structured attack vector for evaluating the system based on these potential attack scenarios. Typically a wireless connection would be made to the AMR device for the purposes of troubleshooting, configuration changes, calibration for correct power reading, and firmware upgrades. However, this capability also provides an opportunity for a potential attacker to discretely drive up next to an AMR residential device and change its operation or to download privileged information. The operations performed by the power company technician would always be consistent, as they would normally run a program that would perform the same operations while performing maintenance. For example, the operations would be:

1. Login
2. Password
3. Perform diagnostics
4. Read calibration values
5. Read phone number/Host IP address
6. Read next connection time
7. Download event log
8. logout

Note that each one of these operations may map to one or more taxonomic events. The events that occur during this process could be monitored in different ways using

Table 1 Candidate taxonomic events

Taxonomic event
200—System started
201—Watchdog timer expired
202—Power consumption sampled
203—Flash memory updated/changed
204—Flash memory cleared
205—Flash memory read
206—Calibration data changed
207—Started power record upload
208—Phone number dialed/connected to host
209—Login sent
210—Password sent
211—Host verified meter
212—Meter verified host
213—Login failed
214—Transfer complete
215—Received new dial-in phone number
216—Received new dial-out time
217—Perform diagnostics
218—Logout
219—Diagnostics pass
220—Diagnostics fail
221—Read calibration values
222—Read phone number/ host IP
223—Read next connection time
224—Download power records
225—Download event log
226— Phone number/host IP address changed
227—Calibration values changed

the SCADA-Hawk system. One way would be instrumented firmware which reports these events when they occur by identifying, through the instrumentation, which parts of the firmware code perform each of the above tasks. The second method of monitoring would be to use the COLLECTORs to monitor the communication commands/packets sent between the wireless device and microprocessor.

In comparison to the above set of sequences, the would-be attacker could tamper with the meter by connecting to it using the same connection point, and using a different sequence of operations. For example, they may issue commands to change the calibration values, phone number, erase recorded power records, etc… As a result, an anomaly would be reported due to the difference in events (Table 2).

To evaluate this particular attack scenario, we encoded such operations as the processing of logging in, password authentication, various diagnostic routines, retrieving calibration values, event log downloading, and other operations into their

Table 2 Normal operations for technician

Taxonomic Event
209—Login sent
210—Password sent
217—Perform diagnostics
221—Read calibration values
222—Read phone number/ host IP
223—Read next connection time
225—Download event log
218—Logout

corresponding ComET events. Table 1 lists the possible monitored events with their corresponding Taxonomic Events. Table 2 identifies the sequence of monitored events observed by the power company technician during the "normal" operation.

In this scenario, the COLLECTORs and AGENTs were configured to have a single AGENT and a COLLECTOR for each event. Thus, the STEB may contain events shown in Table 3, after a normal transmission. The entries are in the following format: Event Time/Taxonomic ID.

The STEB, in Table 4, shows the malicious event occurring in COLLECTOR C5.

The technician may need to change the phone number, host IP, or calibration values. However changing these values may also generate new “patterns” or “sub-patterns” to their operation that can be modeled in the TDM’s. Note that although this scenario is relatively simple, it established a baseline to which we could evaluate the system operations and determine future enhancements. The comprehensiveness of the monitoring mechanisms is only limited by the available system instrumentations for COLLECTORS.

Table 3 STEB Containing normal behavior

C1	C2	C3	C4	C5	C6	C7	C8	C9
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
111/ 209	248/ 210	457/ 217	788/ 221	…	981/ 222	1251/ 223	5891/ 225	9346/ 218

Table 4 STEB Containing abnormal behavior

C1	C2	C3	C4	C5	C6	C7	C8	C9
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
111/ 209	248/ 210	457/ 217	788/ 221	838/ 227	981/ 222	1251/ 223	5891/ 225	9346/ 218

5 Conclusion

To summarize, the evolution and maturation of the smart grid technologies will help to better manage the power delivery systems. However, along with these new technologies will emerge new vulnerabilities that need to be considered for maintaining secure power grid operation. We have presented in this chapter a review of methods of tamper detection by behavior analysis along with a report on our work known as SCADA-Hawk. Our technology consists of passive groups of devices known as COLLECTORs and AGENTs that together monitor device behavior by learning normal behavior and watching for out-of-tolerance anomalies. These devices build learned behavior models by normalizing the SCADA event sequences into a standardized Taxonomy and using our "Snap-Shot" learning algorithms for model building. In addition, we developed an "Event-Player" technology that is used for describing a designated operational event sequences for both normal operation and abnormal operations for evaluating the effectiveness of the SCADA-Hawk system.

Several local SCADA system users are interested in our work and have asked to be kept up to date on our progress. These interests, along with our prototype results, are encouraging and thus we are continuing with the development of SCADA-Hawk. For future work we are looking other types of SCADA equipment, evaluating the types of attack vectors and SCADA events, and other parts of the system that can be improved. In addition, we found that SCADA-Hawk has a unique opportunity for success in infrastructure protection marketplace due to the support of legacy system compatibility.

Several enhancements and refinements to the SCADA-Hawk system's operations have been identified. First, we are considering enhancements for improved noise tolerance of the learning and monitoring algorithms and discriminating the difference between normal and abnormal behavior. Second, to supplement our sequence comparison logic, we plan on integrating the use of statistical data on individual event transitions. In addition, the ability to handle events occurring very close in time requires a very fine granularity clock and decreasing the sample period may be necessary. Furthermore, a means of "triangulation" can be supported due to the system's unique architecture by having multiple COLLECTORs report to a single AGENT in a cluster. This "triangulation" would provide a means for separate event sequences to cross check each other in the regards to a sequence set that commonly occurs together. This would come into play for detecting a root-kit that reports an erroneous system status message at the application layer and cross-checking it against the simultaneous low-level system event sequences that should occur during the application layer. Last, it is planned to convert the COLLECTORs and AGENTs into their respective hardware implementations as the technology matures.

Acknowledgments Our thanks to Air Force Research Labs for awarding us a Phase 1 grant, OSD09-T003, to pursue our research on the SCADA-Hawk project. In addition, we want to thank the power industry experts we met with who helped us understand the industry's security needs and types of events that could be monitored.

References

Cheung S, Dutertre B, Fong M, Lindqvist U, Skinner K, Valdes A (2007) Using model-based intrusion detection for SCADA networks. In: Proceedings of the SCADA security scientific symposium, Florida, Jan 2007

Fang X, Misra S, Xue G,Yang D (2011) Smart grid—the new and improved power grid: a survey. IEEE Commun Surv Tutorials (in press)

Hansen SD (2008) An intrusion detection system for supervisory control and data acquisition systems, Masters by research thesis, Queensland university of technology

Lemos R (2005) U.S. makes securing SCADA systems a priority at http://www.securityfocus.com/news/11351/1

Megalingam RK, Krishnan A, Ranjan BK (2011) Advanced digital smart meter for dynamic billing, tamper detection and consumer awareness. In: Proceedings of the 3rd international conference on electronics computer technology (ICECT (2011), Kanyakumari, India, 8–10 Apr 2011

Oman P, Phillips M (2007) Intrusion detection and event monitoring in SCADA Networks. In: Goetz E, Shenoi S (eds) Critical infrastructure protection. Springer, New York, pp 161–174

OWL 2 (2009) Web Ontology Language Document Overview. World Wide Web Consortium (W3C). Web. 25 Jan 2012. http://www.w3.org/TR/owl2-overview/

Peterson D (2004) Intrusion detection and cyber security monitoring of SCADA and DCS Networks, Presented at ISA AUTOMATION WEST, Copyright 2004 by ISA—The Instrumentation, Systems and Automation Society. http://www.isa.org,citeseerx.ist.psu.edu/viewdoc/download?

Siobhan G (2009) Electricity grid in U.S. penetrated by spies, wall street iournal online, http://online.wsj.com/article/SB123914805204099085.html

Sousan W, Gandhi RA, Zhu Q, Mahoney W (2011) Using anomalous event patterns in control systems for tamper detection, 7th annual cyber security and information intelligence research workshop, CSIIRW,(2011) Oak Ridge National Laboratory. Oak Ridge, TN

Yang D, Usynin A, Hines JW (2006) Anomaly-based intrusion detection for SCADA systems. http://entrac.iaea.org/i-and-c/tm_idaho_2006/cd/IAEA%20Day%202/Hines%20paper.pdf

Zetter K (2012) Hackers breached railway network, disrupted service, Wired, Jan 24, 2012. http://www.wired.com/threatlevel/2012/01/railyway-hack/

Automating Electric Substations Using IEC 61850

Peter J. Hawrylak , Jeyasingam Nivethan and Mauricio Papa

Abstract The Smart Grid will enhance the generation, distribution, transmission, and use of electricity by incorporating elements that will greatly help improve energy efficiency. In addition to traditional components, it will also incorporate small-scale generators, such as home wind turbines and solar panels, into the larger grid. In order to enable energy efficiency as well as other features, two-way communication between utilities and customers (users) will be required. This communication will most likely travel in large part over public networks. The Smart Grid, through the addition of bi-directional communication links throughout the infrastructure, will enable utilities to enhance their service, monitoring, and maintenance activities. Electric power substations will play a major role in the Smart Grid. IEC 61850 is a family of standards that defines network protocols, and data and device naming conventions for electric substation automation. IEC 61850 provides utilities with the ability to better monitor operation and even remotely control the substation when necessary. Part of the utility-substation communication link will be facilitated by public networks (e.g. the Internet). This chapter provides an overview of the IEC 61850 standards and discusses recent experiences with IEC 61850. Challenges facing IEC 61850 deployments, namely security, are presented. Potential solution paths to these challenges are provided.

Keywords IEC 61850 · Smart grid · Security · Automation · Ethernet

P. J. Hawrylak (✉)
Department of Electrical Engineering, University of Tulsa, Tulsa, OK, USA
e-mail: peter-hawrylak@utulsa.edu

J. Nivethan · M. Papa
Tandy School of Computer Science, University of Tulsa, Tulsa, OK, USA

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_6,

1 Introduction

Modern electric distribution and transmission systems are complex and must respond to varying customer consumption patterns. The purpose of the transmission system is to transfer the electricity generated at the power plants to the customer (end user). Substations are positioned between the power plant and the customer to help distribute electric energy. The power plant-to-substation link is typically constructed using very high voltage lines in order to minimize the current flowing through the lines, thus, reducing power loss due to the resistance of the lines. Customers require low-voltage power, on the order of 120 VAC at 60 Hz in the United States, and the substation provides the conversion from high-voltage to low-voltage that are more suitable for final distribution to customers. In larger distribution networks there may be a series of substations, each stepping the high-voltage output of the power plant to a lower voltage for transmission to the next substation. Ultimately, a substation provides electricity to a group of customers.

Substations are often located in remote, rural, or difficult to access locations because electricity must be provided to each customer at the location where it is needed. Therefore, the cost of sending an engineer or technician to a substation can be significant and utilities would like to reduce the number of times an engineer or technician must visit a substation. In addition, this service model prevents utilities from making real-time changes to the operation of the substation. Inability to make real-time operational adjustments prevents the substation, and by extension, the electric grid from instantly responding to changes in electric demand. Real-time control and updates of device operating settings to optimize electricity delivery and use are two of the major benefits of the next-generation electric grid, termed the *Smart Grid*.

Previously electricity was provided to the electric grid only by power plants operated by electric utilities and flowed in one direction: power plant to end-user. Communications throughout the grid, for the most part, were also unidirectional in nature. With the advent of small-scale generation, such as solar panels and small wind turbines, end-users can generate their own electricity. The Smart Grid will enable end-users to sell excess energy back to the utility. This requires an electric grid and control system that supports a bidirectional flow of electricity between utility and end-user. To determine how to handle this bidirectional flow of electricity, the Smart Grid must be informed when users will push electricity onto the grid and this requires bidirectional communication between the utility and the end-users. The Smart Grid will utilize bidirectional communication links between end-user and utility to manage and control the flow of electricity. Many different types of communication technologies will be used to construct the communication infrastructure of the Smart Grid. An overview of the communication technologies that could be used in the Smart Grid is provided in Wang et al. (2011). Substations will play a key role in the Smart Grid by adjusting the amount of electric power they provide in response to current demands and the amount of customer generation (e.g. excess power from solar panels) provided to the grid. These types of functional requirements will in turn require the substation and substation equipment to be interconnected to

coordinate action for normal substation operation, receiving updates from the utility, and receiving information from end-users.

Substations in the Smart Grid will employ local area networks (LANs) to connect internal devices together to reduce construction and management costs. The Smart Grid offers utilities an increased level of control and remote operation of substations. The shift from direct point-to-point connections to a LAN environment requires reevaluation of the operating and security guidelines and best practices. Remotely controlling substations requires enhanced safety systems that can take over control during faults or unknown conditions to maintain safe operation of the substation and maintain the Smart Grid (Myrda and Donahoe 2007).

The Smart Grid will consist of a number of different network technologies and devices. Interoperability is a key requirement for the Smart Grid. Interoperability can be provided by standardizing how data is presented and communicated. Efforts in this area are underway within the IEEE, US National Institute of Standards and Technology, and the International Electrotechnical Commission (IEC). An overview of the various standards being developed within these groups is provided in Wang et al. (2011).

The IEC 61850 family of standards provides guidelines to implement communication within the substation and between the electric utility and substation. Thus, the IEC 61850 standards enable the automation of substation control from a central location. This automation provides the ability for the electric utility to respond to changing conditions in real-time and achieve one of the primary requirements for the Smart Grid. However, this automation or remote control introduces a number of issues that must be addressed in order to provide a safe and reliable Smart Grid. This chapter explores these issues. First, an overview of the IEC 61850 family of standards is presented, followed by a discussion on how substations can be automated using the IEC 61850. Next, the challenges in achieving and maintaining the remote control substation are discussed and potential solution paths to address these challenges are presented. The chapter concludes with a presentation of the necessary future work and research directions in this area.

2 Brief Overview of IEC 61850 Standards

The IEC 61850 standard is divided into ten parts with some parts having multiple sub-parts. A number of other technical reports are included with these parts. These technical reports provide additional information about the application of IEC 61850 and some include descriptions of IEC 61850 use-cases or potential implementations. Brief summaries of the technical reports follow the overview of the standards.

IEC 61850-1 provides an introduction and overview to the IEC 61850 family of standards. The introduction covers basic information about protection, control, and monitoring needs for electric; it is not a comprehensive overview of electric protection, control, and monitoring. The goal of this part is to introduce the fundamentals of IEC 61850 and provide illustrative diagrams and figures.

IEC 61850-2 contains the glossary of terms used in IEC 61850. This document houses these definitions for the entire family of standards and is updated as needed when other parts are updated or modified.

IEC 61850-3 contains the basic operating requirements for the IEC 61850 network. These include quality, security, environmental conditions requirements, and voltage requirements for the communication network.

IEC 61850-4 describes the system and project management aspects of IEC 61850. This includes quality assurance, beginning with the design stage, and engineering requirements. This part provides management guidelines for the entire life cycle of the system from design through decommissioning.

IEC 61850-5 explains and identifies the communication requirements of the all known tasks (functions) performed in the substation automation system (SAS). This part covers the lower level of the communication stack and is intended to be flexible to accommodate a range of devices and expandable to adapt to new networking technologies. Descriptions of the logical node concept and performance requirements for the network are also provided.

IEC 61850-6 describes the substation configuration language (SCL) that is used to initialize and describe devices within the substation. The SCL is used to describe the communications between devices, and the relationship between devices and functions performed by the substation.

IEC 61850-7-1 explains the basic communication architecture among the devices. It provides the communication modeling methods, modeling concepts, communication principles and the information models discussed in IEC 61850-7-x documents. Communication modeling methods, modeling concepts, communication principles and the information models used in the remaining subparts of IEC 61850-7 are described. Data formatting information is provided to enable interoperability between devices.

IEC 61850-7-2 describes and explains the abstract communication service interface (ACSI) that is used for the substation commination in the IEC 61850 model. ACSI is defined to be protocol agnostic and provides the basis to build the IEC 61850 protocol. This abstract interface provides services for (1) communication between the client and a remote device; (2) the transmission of messages using the publisher/subscriber method; and (3) the transmission of sampled values using the publisher/subscriber method. In the publisher/subscriber method, the each device monitors the network for messages to which it subscribes (that are of interest). Each message contains publisher and identification information so that the subscribers can identify those messages as being of interest.

IEC 61850-7-3 specifies common data classes and common attribute types for substation applications. The data classes and common attributes are used to define communication between applications.

IEC 61850-7-4 defines naming conventions for logical nodes and data structures used in the IEC 61850 model. The standardized naming convention is central to providing interoperability between devices and enabling information sharing.

Several standards and technical reports (denoted by a TR in the IEC standard name) have been published or are under development (IEC 61850-7-x10 and IEC

61850-7-x20) to provide extensions of the IEC 61850 naming and data format conventions to applications outside of the substation. Extensions to distributed energy resources (IEC 61850-7-420) and hydroelectric facilities (IEC 61850-7-410 and IEC/TR 61850-7-510) have been published.

IEC 61850-8-1 defines and explains a method for mapping the ACSI messages to Manufacturing Message Specification (MMS) messages in order to transfer non-time critical data through local area networks, and a method for mapping ACSI to Ethernet in order to transfer time critical data through local area networks.

IEC 61850-9-1 has been withdrawn and was replaced by IEC 61850-9-2. IEC 61850-9-2 defines and explains the specific communication service mapping (SCSM) of sample values to Ethernet. This standard describes how to take the ACSI model defined in IEC 61850-7-2 and convert those data items for transmission over Ethernet.

IEC 61850-10 defines the conformance requirements for IEC 61850 installations. This standard defines the tests and pass/fail conditions for determining conformance to IEC 61850.

Technical reports having the numbers IEC/TR 61850-90-x describe how to use IEC 61850 and provide additional information about implementation. Two technical reports have been published, IEC/TR 61850-90-1 covering communications between substations, and IEC/TR 61850-90-5 covering how to transmit synchrophaser data using the IEEE C37.118 standard.

3 Substation Automation with IEC 61850

The IEC 61850 series of standards defines how data is represented, the network protocols for transmitting the data, and the network requirements for substation automation. The benefit of IEC 61850 is that it helps replace the dedicated connections to each device with a single Ethernet connection. This provides significant savings in cable cost and more sophisticated functionality provided by more advanced devices capable of interacting with any other device in the substation.

The communication infrastructure utilized by the current power grid is composed of dedicated point-to-point connections between a local controller and the device. Networks based on this type of topology do not scale well and are limited to small geographic areas (Wang et al. 2011). An IEC 61850 based Smart Grid will use a different topology that is similar to what is often seen in current LAN designs and deployments, most likely based on Ethernet. In fact, current versions of IEC 61850 use Ethernet (Ozansoy et al. 2009) as the network infrastructure although the protocols defined by IEC 61850 could be transported using a number of other network infrastructures. Such a network can be connected into the same network that is used for providing Internet connections to business and private homes.

IEC 61850 defines the network protocol for intelligent electronic devices (IEDs). IEC 61850 focuses on Layers 5 (session), 6 (presentation), and 7 (application) of the OSI (Open Systems Interconnection) model (Forouzan 2003). One of the goals of IEC 61850 is to provide a mechanism to describe data and control commands

independently from the underlying network (Mackiewicz 2011). To accomplish this, IEC 61850 defines the data format for the messages and the data labels in each message. Each message is constructed at the higher levels in the IEC 61850 format and is then translated and encapsulated by the network infrastructure into the appropriate format for transmission. Upon reception, the message is again translated by the network infrastructure and the receiver processes the resulting data according to the IEC 61850 standard. This allows the IEC 61850 protocol to be used with a number of network technologies (e.g. Ethernet, WiFi, or ZigBee). Being network agnostic is important because it enables the IEC 61850 protocol to continue to be used as network technologies advance.

The Smart Grid also requires devices to communicate with devices from other vendors and this requires interoperability, which is a major goal of IEC 61850. The issue of interoperability is addressed by defining the structure and ordering of data items in messages and how these data are presented in those messages (Mackiewicz 2011; Xiong et al. 2008). The goal is to achieve plug-and-play capability (similar to the capability provided by USB for PCs) for IEDs to provide the interoperability necessary for the Smart Grid (Hossenlopp 2007). In the Smart Grid, the equivalent of the driver for the hardware device that plugs into the USB port is the IEC 61850 definition of the data items and how those data items are presented. Standardizing the data and their presentation also reduces the amount of work required to connect IEDs to the Smart Grid network because each device uses the same standardized communication protocol and data definitions (Mackiewicz 2011).

The base component of an IEC 61850 network is the physical device, e.g. a smart relay. The physical device has a defined location and purpose and is given a network address (Mackiewicz 2011). A physical device may perform a single function or multiple functions, e.g. measuring current, status monitoring, or controlling a circuit. Each function that a physical device performs is considered a logical device (Mackiewicz 2011), and is similar to the concept of a port in computer networking. Hence, a single physical device may contain multiple logical devices: one for each function. In turn, each logical device is composed of multiple logical nodes, which are groups of data items and procedures provided by the logical device (Mackiewicz 2011). The data items can be measurements or set-point values, while the procedures perform the actions required to maintain the power grid. The format and presentation of these data items is provided in IEC 61850-7 (Mackiewicz 2011). The architecture of IEC 61850 lends itself well to being implemented in Object-Orientated code. In Ozansoy et al. (2009), the authors describe an Object-Oriented framework that can be used to implement IEC 61850 IEDs and the communication exchanges between those IEDs. In Mercurio et al. (2009), the authors present an implementation in simulation of a substation communication and control network based on the IEC 61850 and IEC 61970 standards using the Object-Oriented program design philosophy. They define a framework for building software objects to represent the components of a substation network and investigated the use of the common object request broker architecture (CORBA) to send messages (Mercurio et al. 2009). The use of CORBA does not allow this implementation to be used in real-time monitoring or control systems that have hard deadlines (Mercurio et al. 2009). Although use of CORBA has

declined significantly over the last few years, integration efforts within the CORBA framework can be easily ported to other frameworks.

3.1 IEC 61850 Network Architecture

IEC 61850 defines two different buses or local networks: the station bus and the process bus (Mackiewicz 2011). The process bus connects the relays and primary equipment to the IEDs in the substation (Mackiewicz 2011). The process bus is typically deployed at the bay level and may cover a single bay or span multiple bays. Data aggregators termed merging units combine information from multiple IEDs into a single message and then pass this on to the process bus (Mackiewicz 2011). This helps reduce traffic on the process bus and reduces the number of network connections the process bus must support. The IEDs act as the bridge between the process bus and the station bus. The station bus connects all of the process buses together and provides the connection to the utility control center (outside world) and other substation management functions (e.g. substation historian) (Mackiewicz 2011). Figure 1 illustrates the process and station buses in relation to substation management and primary equipment.

An alternative approach to having both a station and process bus is to implement only a station bus. In this case, hardwired direct connections are still present between the IEDs and the relays and primary equipment. The former option (station and

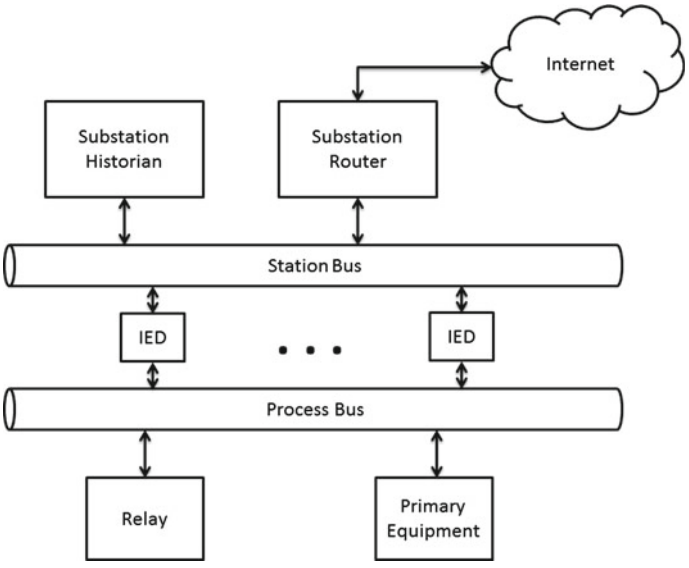

Fig. 1 Typical IEC 61850 network architecture with process and station buses

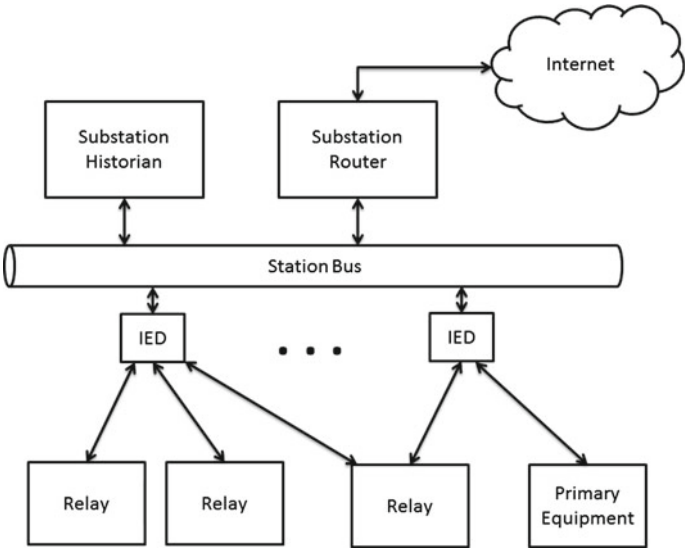

Fig. 2 Typical IEC 61850 network architecture with only a station bus

process bus) is often employed for new substations and this option (station bus only) is useful for a retrofit of an existing substation to upgrade it to use IEC 61850. This option minimizes the impact on the internals of each bay because only the connection between the station bus and bay equipment must be updated. Figure 2 illustrates the basic network architecture for the station bus only option.

IEC 61850 defines two communication functions to provide an interface between an IEC 61850 IED and a supervisory control and data acquisition (SCADA) system (Higgins et al. 2011). Both functions deal with the IED reporting changes or alerts in signals that the IED is monitoring to the SCADA controller. Alerts can be generated when a signal is detected outside of a predefined range (e.g. exceeds a threshold value) or when a signal changes too quickly (Higgins et al. 2011).

Two types of connection strategies are possible: data gathering only, or data gathering and control. Initial implementations were data gathering only and are not required to meet the timing requirements of control systems. In these systems, the data gathering connections are simplified into a single Ethernet connection that links the device back to the central control center.

Control applications require time critical operation. Typically, faults must be identified, processed, and corrected within two cycles. IEC 61850 specifies that time-critical messages must be delivered within 4ms from the time they are sent. The 4ms requirement in less than the two-cycle time but it is important that the message arrives in time for the protection device to process, respond and take action. Thus, the IEC 61850 network must provide the communication bandwidth and transit time to meet this requirement.

Support vector machines (SVMs) (Cortes and Vapnik 1995) are used in pattern matching applications and are one technique to monitor power systems for stability. SVMs can also be used to help secure power systems by identifying abnormal behavior patterns, and can be integrated into an intrusion detection system (IDS) for the power sector. SVMs must be trained with known input sets representing normal and abnormal behavior or conditions. Once trained, SVMs can be used to identify suspicious behavior in the Smart Grid. SVMs have been used to forecast electricity demand based on data from Smart Meters (Men and Liu 2011).

IEC 61850 provides support to transmit sampled value data, which are data taken from sensors. Merging units can be employed to collect and aggregate sample value points together to send in one large packet verses several smaller packets, reducing network load in terms of the number of messages. Thus, the IEC 61850 sampled value message type can be used as input to SVMs. The critical concern is for the network and system to be capable of providing data input at the required rate.

3.2 IEC 61850 Deployments and Pilot Projects

The Tennessee Valley Authority (TVA) designed their Bradley Substation to use IEC 61850 (Ingram and Ehlers 2007). IEC 61850 provides a substation configuration language that can be used to encode the equipment description for an IEC 61850 network. Naming conventions for the SCL description are defined in IEC 61850 to simplify identification of information to other IEDs and devices (Ozansoy et al. 2009). In this manner, the SCL provides part of the driver for the plug-and-play capabilities of the IED. The SCL also provides the ability to describe the connections between devices in the substation (Higgins et al. 2011). Thus, the SCL maintains the network topology for the substation. This enables the user to obtain a snapshot of the network topology at any time. Using the SCL provides automated book-keeping for new devices or those that are replaced: changes in the SCL can be easily documented by periodically requesting the SCL information from the IEC 61850 network. This is very important for auditing and scheduling of routine maintenance of equipment. Further, each piece of equipment can be tracked even if it is moved around in the substation or to a different substation. Such tracking is important because it allows detailed maintenance and service logs to be kept for each device. TVA cited the SCL as one of the major benefits of IEC 61850 because of the labor savings in IED setup and installation, and in the simplification of device management in the future (Ingram and Ehlers 2007; Rietmann and Reinhardt 2006).

TVA will realize a significant savings due to automation because routine inspections of the substation can be limited. This automation requires that TVA be able to remotely query IEDs to see not only their published readings but also internal settings and measurements to enable TVA to remotely diagnose problems (Ingram and Ehlers 2007). Remotely diagnosing problems will ensure that the technician sent to repair the substation has the proper training and equipment (Ingram and Ehlers

2007). This will reduce operating expenses, as technicians know exactly what the problem is and already have a plan to address it before arriving at the substation.

The Comisión Federal de Electricidad (CFE) upgraded some of their substations to an IEC 61850 based network beginning in 2005 (Moreno et al. 2012). They cite interoperability, interchangeability—the ability to swap devices with ones from different manufacturers, use of an open and royalty free protocol, vendor independence—not being restricted to a single vendor's vertical solution, and a LAN network to link components together in a cost-efficient and safe manner as the primary goals/benefits of moving to IEC 61850 (Moreno et al. 2012). They found that documentation of the network configuration, especially, maintaining a current list of the IP address assigned to each IED was critical (Moreno et al. 2012). Interchangeability was possible with only one manufacturer's IED and that IED stored the SCL internally (Moreno et al. 2012). Hence, for interchangeability, it is recommended that IEDs store the SCL internally.

It is important to note that CFE found that IEC 61850 did not provide all of the alarm and data types they required, but they were able to extend the IEC 61850 data types to accommodate their new requirements (Moreno et al. 2012).

4 Challenges for Substation Automation

There are a number of challenges to using IEC 61850 for substation automation. Moving to a LAN structure introduces many variables in the communication infrastructure that must be taken into account. Maintaining accurate date and time settings and maintaining quality of service under variable traffic loads are two examples. The non-deterministic nature of an Ethernet network complicates these requirements and introduces other issues. In particular, the non-deterministic nature of Ethernet may produce scenarios where packets containing measurements arrive out of order. This is due to network relay devices (e.g. switches and hubs) that must retransmit packets as they move between networks. Large packets may be broken up into smaller packets based on network infrastructure or current traffic loads. The IEC 61850 network must ensure that all parts of the message reach the destination and are reassembled within the 4ms time window. Using commercial networks to link the substation to the utility as opposed to dedicated private connections, requires strong security on both the IEC 61850 network and the physical devices themselves. These challenges are described in this section and potential solutions to each challenge are described.

4.1 Timing Synchronization

In a distributed control system, it is important that the IEDs are synchronized with each other. This enables accurate timestamps to be applied to data readings that are used by data aggregators to make control decisions. This is especially impor-

tant with Ethernet because messages may arrive out of order: readings at time t+30 may arrive before readings from time t. There are many methods of clock synchronization available and one is the network time protocol (NTP) that is standard on most Ethernet TPC/IP based networks. The IEC 60870 family of protocol defines the communication between the substation and control center (Sanchez et al. 2010). In Sanchez et al. (2010), the authors present an implementation and simulation of an IEC 60870-5-104 protocol using standard Internet protocols, SNMPv3 (simple network management protocol ver. 3), SSH (Secure SHell), and network time protocol (NTP), instead of the IEC 60870-5-104 stack. They use a database table system, called management information base (MIB) (Stallings 2000), to provide read and write access to the intelligent electronic devices (IEDs) (Sanchez et al. 2010). The MIB concept is widely used in networks to manage the devices connected to the network (Stallings 2000). They compare their implementation to an implementation using the IEC 60870-5-104 stack in terms of number of bits transmitted and percentage of communication that is related to electric power data and control. The method proposed by Sanchez, et al., yields better performance than the IEC 60870-5-104 stack: fewer bits transmitted and a higher percentage of messages containing useable information verse book-keeping information (Sanchez et al. 2010). However, using NTP they are only able to obtain a synchronization within 10 ms (Sanchez et al. 2010). Hence, their proposed system is not suitable for control and monitoring applications requiring hard real-time deadlines of under 10ms and is not suitable for IEC 61850. Other synchronization methods, such as IEEE 1588 (IEEE 2008) and GPS (global positioning system) timing information, may yield better results. Variable network traffic must be taken into account and proposed solutions to the synchronization and other problems must be analyzed to determine their impact on the overall network load and message transit times.

Clock synchronization is important and this synchronization will require additional traffic on the network. Other maintenance tasks will add to this overhead. One option is to use a separate (redundant) network to provide synchronization and other maintenance tasks, leaving the primary network to carry just monitoring data and control information. This has the added advantage of providing another redundant connection that could be used for other forms of data transfer during an emergency.

4.2 Network Traffic Load

The amount of messages present on the network is directly related to the level of congestion on the network. Gao, et al., investigate the congestion of a network based on the amount of bandwidth that is utilized on the network (Gao et al. 2008). They found that under normal conditions that messages could be delivered within the 4 ms limit, but that a slight increase in utilization caused message delays to exceed the 4 ms limit (Gao et al. 2008). Periodic status reports and heartbeat signals from physical devices can be scheduled to avoid such congestion. However, during a fault or emergency condition the network utilization will increase substantially leading

to increased congestion and delay. One alternative is to provide multiple network links to each device, with each link serving a different type of message class. For example, one link could be dedicated to periodic status and heartbeat signals, while a second could be dedicated for control and emergency messages. Unfortunately, the use of multiple links is contrary to the goal of IEC 61850 to reduce the number of communication links in a substation. Another solution is to increase the bandwidth of the network infrastructure (Gao et al. 2008). Increasing the bandwidth solves the present problem, but most likely only for a short time because the extra bandwidth will be used up to provide new features or improved control.

The architecture of the Ethernet network is important to its performance. The network must be designed to minimize packet delay in order to meet the 4ms timing requirement specified by IEC 61850. In their initial deployment, CFE used two ring topologies connected at a few places for their substation network (Moreno et al. 2012). This architecture did not perform well and resulted in significant delays during fault conditions (Moreno et al. 2012). A single ring topology provided better performance when clearing a fault (Moreno et al. 2012). As a result, CFE's recommended network topology is a single ring with one Ethernet switch per panel (Moreno et al. 2012). Using one Ethernet switch per panel will dramatically reduce the number of connections to the backbone substation network and allow intra-panel messages to be delivered without having to travel over the backbone substation network. This provides significant reduction in network traffic, thus, improving performance on the process bus.

4.3 Communication Protocol with Non-Deterministic Timing

The Ethernet protocol does not provide deterministic transit times for messages from sender to receiver. First, Ethernet defines a non-deterministic exponential back-off time when a collision or consecutive collisions are detected. While this method works well to address collisions, it yields a protocol where the delay in transmitting a message from the sender to the physical network varies with current network traffic. Second, Ethernet networks are multiple hop networks with intermediate devices, e.g. hubs, switches, and routers, routing messages to their proper destination. As these intermediate devices receive input from multiple senders, incoming messages are placed in a queue and are processed sequentially. When sending the message the first issue is again present. This issue is also dependent on network traffic, but also on network topology. Network topology is important because it defines choke-points, e.g. a central switch, in the system that receive an increased amount of traffic. Network topologies for IEC 61850 substations must take this into account and not route too much traffic through a single choke-point (e.g. switch).

Evaluations (Kanabar and Sidhu 2011) have shown that IEC 61850 can provide the needed latency for data reporting, but not for control. Control requires at least two messages to be sent. The first message is from the device to the central controller reporting a fault or abnormal data. The central controller then responds with a mes-

sage indicating what action to be taken to clear the fault. Possible solutions to the latency problem are described below.

High-end Ethernet switches have Quality of Service (QoS) capabilities that allows system designers and system administrators to attach a priority to a message or communication channel. Higher priority messages are given preference over lower priority messages and are relayed first. Sidhu and Yin investigate the performance improvement of QoS using a network simulation of a sample IEC 61850 enabled substation (Sidhu and Yin 2007). Their results indicate that using QoS improves performance (lower transit time) for an Ethernet network based on 10Mb/s equipment, but that no improvement was observed for an Ethernet network based on 100 Mb/s equipment (Sidhu and Yin 2007). Thus, the QoS feature may not provide the needed savings to meet the timing requirements of substation for higher speed networks.

As the number of devices connected to the network increases so will the transit time of messages due to increased contention and traffic being routed though choke points. Scalability was also found to be an issue as a 3.14X increase in transit time was observed when the number of Ethernet switches in the example substation was increased to 14 from 9 (Sidhu and Yin 2007). Detailed network simulations must be conducted to verify that the proposed network topology for a substation can provide the required speed. Choke-points must be identified and the topology changed as needed to provide the required QoS.

The bandwidth of the network is an important consideration and must be sufficient to handle the network traffic while ensuring delivery of messages within the required time. Messages in the substation will have different priorities based on type and current operating conditions. Message priority can change as the current operating conditions change making it difficult to meet the timing requirements using the QoS features of Ethernet (Deshpande et al. 2011). Deshpande et al. (2011) propose a strict priority queue (SPQ) and provide a priority ordering of Smart Grid message types for the Differentiated Service Code Point (DSCP) feature in the IP header. The SPQ feature provides a means to enable high priority messages to meet their delivery time requirement by preempting lower priority messages that have larger maximum allowable delays (Deshpande et al. 2011). They simulate a small substation network to determine the bandwidth required to meet the message delivery time requirements. In their simulation the maximum allowed time for a message to be delivered is 8 ms, which is twice that allowed by IEC 61850 (Deshpande et al. 2011). Their results show that a network bandwidth of 3.072 Mb/s is required to meet the 8 ms delivery requirement (Deshpande et al. 2011). This translates into a bandwidth requirement of 7.144 Mb/s for an IEC 61850 network. This bandwidth is easily achieved using a standard 10 Mb/s network. Most Ethernet installations today offer a bandwidth of 100 Mb/s so this should easily meet the message delivery requirements. However, differences in the traffic on the network and network architecture from the simulated network will alter the results. Thus, a higher or lower bandwidth may be required for an arbitrary network.

The use of multiple links in a substation can be used to reduce message transit time. One method of reducing the transit time is to use these multiple links in the same manner as a parallel I/O port. Large messages can be broken into smaller parts

and transmitted over several links in parallel (Deshpande et al. 2011). The Multi-link Point-to-Point Protocol (MLPPP), defined by IETF (Internet Engineering Task Force) RFC (request for comments) 1990, can be used to provide this functionality (Deshpande et al. 2011). The use of MLPPP may result in a lower bandwidth requirement (Deshpande et al. 2011). The redundant links in the substation could be used to provide the parallel connections. However, this use of the backup (redundant) network links must be weighed against the need to have backup network links in the event that some of the primary links fail. Network bandwidth should not be designed using MLPPP as some of the parallel links are provided by the backup network in the substation. The required bandwidth must be computed based only on the non-backup links and enough redundant links must be installed to provide the MLPPP service.

4.4 Order of Packet Reception

The Ethernet protocol does not guarantee that the receiver will receive the packets in the same order as the transmitter sent them. Higher level protocols, such as TCP guarantee in-order delivery by default and IP provides means to establish packet order at the receiver. This can be problematic for data monitoring applications that are performing calculations to monitor the stability of a power line in real-time. IEC 61850 uses a mix of protocols (e.g. Ethernet, IP, and TCP) transmitted over an Ethernet medium. The GOOSE (Generic Object Orientated Substation Event) IEC 61850 message is transmitted at the Ethernet level and the Ethernet header does not contain any information about ordering. Hence, the packet must include a field to denote the order or a timestamp having the required precision associated with each reading. This is necessary to allow the receiver to reconstruct the ordering of packets and to process the data in the appropriate order. This added field increases the time required for communicating the message (more bits to send and receive) and the processing time on each end (receiver must reorder packets). This field may be in the header for the protocol (e.g. fragmentation information in the IP header), associated with an acknowledgement and transmission scheme inherent to the protocol (e.g. acknowledge/retransmit scheme for transmission of data segments using TCP), or as part of the data itself.

The data analysis algorithms used in substations automated using IEC 61850 must support reception of packets in a different order than they were sent by the transmitter. Reordering of messages at the receiver is not always feasible, for example, in monitoring applications where data must be processed to derive current values. In extreme cases, earlier packets may be significantly delayed and arrive at the receiver too late to be of any value or the packet may be dropped (never arrives). Hence, the monitoring algorithms must be designed to account for missing or late data due to the non-deterministic nature of the Ethernet communication medium. Kanabar and Sidhu (2011) present an algorithm to estimate missing data values in their monitoring functions. Based on their results, monitoring systems with higher sampling rates were more resistant to consecutive missing samples.

Estimation of missing data also affects the response time of an IED. Kanabar et al. (2011) investigate the response time of IEDs for monitoring and controlling phase delay in the electric grid. The IEDs they used in their experiments assumed a value of 0 for missing phase measurements (Kanabar et al. 2011). They propose an algorithm to estimate the missing values and their results indicate that the response time of the IED is reduced to 23–25 ms from 23–41 ms without the estimation (Kanabar et al. 2011). IEDs must support robust estimation algorithms to estimate the value of missing data items. These algorithms must provide high quality estimates while also being able to detect fault and abnormal conditions to which the IED must respond. Designers must thoroughly understand how the IED handles missing data values and design the control algorithm and system to handle these events safely.

4.5 Fragmentation of Large Packets

The TCP/IP protocol supports *fragmentation* or the breaking up of large packets into smaller packets (Forouzan 2003). The maximum size of a packet is defined by a number of factors including network capabilities and network traffic. In an IEC 61850 Ethernet based substation, the network is defined to meet a specified criteria and should be uniform across the substation. Hence, network capabilities should not have much impact on variations in maximum packet size. However, in substations that utilize multiple types of physical networks, e.g. Ethernet, WiFi, and ZigBee, the network capabilities may influence the maximum packet size. This is especially true at the interfaces between networks. The current amount of traffic on a network is the major factor in determining the maximum packet length. Longer packets take up space on the medium and other messages may not arrive in time. Smaller packet sizes may be used to send parts of multiple messages in the same time. Here, fragmentation enforces fairness on the communication medium and prevents any one device from being starved out of communication. Message size is an important factor in the time required for a message to be received. One simulation, showed that message sizes of 2 kB were able to be transmitted within the 4ms limit but those of 4 kB were not (Gao et al. 2008). Fragmentation can reduce the delay in transit across the network, but the key metric is not the delay of each fragment (part of the message), but the reception of the entire message (Gao et al. 2008).

One solution is to keep the size of each message small enough to fit into the smallest packet size supported by a physical layer. This will reduce the delay significantly and can allow the entire message to be received within the 4 ms required by IEC 61850. However, sending multiple small messages will increase congestion on the network, introducing additional delay. Thus, the optimal size of the message to reduce transit delay is dependent on the network protocol and the network architecture. The optimal values for message size can be investigated using standard discrete event simulation models of Ethernet. Such simulations should be conducted for each variation of network architecture early in the project.

4.6 Security Implications of Wireless for IEC 61850 Networks

Wireless networks offer many benefits over wired networks. First, the need for cabling between nodes is eliminated. Second, nodes can be located in locations that are difficult to access with a wired connection. For example, a wireless node located on a piece of moving equipment can be placed without the need to provide a connection to it through stationary parts of the equipment. Another example is to place a node in a hazardous environment and communicate using wireless means. Third, wireless networks are more easily upgraded. Additional hardware is not required to add more connections. Although access time will increase as more connections are made, there is no problem of running out of ports as on an Ethernet switch. Finally, a wireless network provides easy access for the technician: through a cellular link or by driving by the substation and connecting with a traditional wireless, e.g. WiFi, link.

Wireless networks are visible to anyone within range with the appropriate receiver. This availability reduces the effectiveness of physical security solutions in protecting the substation. Because an economical infrastructure is one of the goals of the Smart Grid, these networks will be commercially available technologies, e.g. WiFi or ZigBee. While security can be activated on wireless networks this comes at the price of added overhead and slower communication speeds. Further, as computing technology advances, these security measures will become outdated and are subject to being hacked.

Wireless communication can also be affected by electro-magnetic interference (EMI) generated from other wireless equipment or from electric equipment in the substation (Yu and Johnson 2011). EMI can cause messages to be corrupted resulting in a retransmission or not being received at all. Intentional EMI generated from an attacker using a jamming device could be used to disable parts of the IEC 61850 network. Physical security should address the jamming attack by keeping the attacker out of range of the substation. Verification and testing of wireless equipment for the substation for EMI immunity must be conducted and the design engineer must be aware of other devices that will also use the wireless spectrum (Yu and Johnson 2011). Transmission algorithms must be developed to coordinate transmissions among multiple wireless devices. Definition and use of accepted testing standards and best practices will address the EMI immunity concerns (Yu and Johnson 2011).

4.7 Integration of Control and Protection Functions into a Single Device

Older substations have little or no integration of functions for control and protection, typically employing a one function per device model. The logical node feature of IEC 61850 provides the ability to integrate several functions into a single IED. Integration reduces the number of point-to-point connections and the amount of cabling, but increases the impact of a device failure. The IEC 61850 protocol allows

the integrated node to collect and package input from multiple sensors or distribute commands to multiple actuators. Hence, the integrated node can be visualized as a collection of sensors and actuators that communicate with the outside world through a centralized router block.

From a security standpoint, a system composed of a number of individual nodes may be more secure than one with a few nodes, each containing several functions. In the event that a node is compromised in the one function per node case, the attacker can control or affect only one function. However, in the integrated case, the attacker can control or affect multiple functions. If the attacker is able to compromise or disable the centralized router or switch they can block all communication with the IED resulting in the loss of information and capability provided by that IED.

Normal failure of IEDs is another event that must be guarded against. Simple, but robust solutions employ redundant sets of IEDs throughout the substation. With the merging of functionality of several non-IEC 61850 IEDs into one IEC 61850 IED, this becomes a larger concern because the failure of one IEC 61850 IED disables several functions. One method proposed in Xiong et al. (2008) is a generic control IED that sits in the background and only becomes active when there is a fault that disables an IED. This backup IED then provides the analysis and reporting features, filling in for the disabled IED (Xiong et al. 2008). If the disabled IED is in direct control of electric machinery, the spare IED can only provide analysis and reporting information to the central controller. However, if the IEDs directly controlling the machinery are intact the spare IED can establish communication with these IEDs and provide control and well as monitoring functions.

Integration of control and protection functions into a single IED requires a reevaluation of the accepted best practices for providing adequate backup and redundancy for substations. Security is also an issue and best practices need to be developed for security IEDs integrating multiple functions.

4.8 Potential Attack Vectors to the IEC 61850 Network

The IEC 61850 network architecture is susceptible to the same attack vectors that a traditional data network is. The US National Institute of Standards and Technology (NIST) has published a report, NIST Special Publication 800-82 (Stouffer et al. 2011), covering security issues in SCADA and control systems for the critical infrastructure. This report provides an overview of the entry points for threats and attacks into these systems. Recommendations are provided to address and guard these entry points.

The substation network can be divided into two areas or sub-networks. The first sub-network is a local area network (LAN) entirely within the substation and connects the protection devices and monitoring systems together. This network is based on IEC 61850 and is typically contained entirely within the substation perimeter. The substation perimeter provides protection to the substation and IEC 61850 network from physical attacks. The second sub-network is the connection between the substation and the utility. It is more economical to have the substation tie into a

large commercially available network to link back to the utility than for the utility to construct a private dedicated network connection (Dondossola et al. 2009). The commercial network provides the attacker with a potential entry point into both the substation and utility. Often multiple independent connections between the substation and utility exist to address the redundancy requirements of the electrical grid (Dondossola et al. 2009). Each of these links is a target for an attacker and must be secured and monitored. Within the Smart Grid data integrity and device availability are the two key requirements (Giani et al. 2008). Integrity includes verifying that the sender is a legitimate part of the network and not a rouge device (Giani et al. 2008). Availability is critical because the IEDs must respond to faults and changing conditions within a hard-real-time deadline. Potential attack vectors are described in the following sub-sections.

4.8.1 Eavesdropping on Network Traffic

The attacker must acquire knowledge of the system they are planning to attack to identify vulnerabilities. The network architecture is a valuable piece of information that the attacker can use in their attack. Based on this information they can identify the structure of the network, including central Ethernet switches, and can discover the type of devices on the network. Using this information the attacker can select targets on the network that they can exploit.

This information can be obtained through eavesdropping on the messages carried by the network. There are several tools available for sniffing network traffic or the attacker can compromise a device on the network and use that device to eavesdrop. Encryption is one method to counter eavesdropping because the attacker is not able to decrypt the messages. Jung et al. (2008) propose deploying a security device between each device (e.g. a RTU) and the centralized SCADA network to counter the eavesdropping attack. While encryption will counter eavesdropping it will not counter a replay attack were an attacker sends a previously recorded message, e.g. a trip command, at a later time. A replay attack compromises data integrity by reusing a previous message or command to affect some change to the system. It is important to ensure that any solutions used by IEC 61850 include protection against replay attacks.

4.8.2 Countermeasures to the Replay Attack to Ensure Data Integrity

The use of a timestamp in each message or a challenge-response protocol will counter the replay attack. Any encryption or challenge-response based security measure must be implemented to meet the timing requirements of the system (4 ms in IEC 61850). Challenge-response based approaches while powerful from a data security standpoint require multiple messages to be exchanged and this may be problematic from a timing standpoint. An implementation of the encryption/decryption in hardware, such as an FPGA (field programmable gate array), will most likely be necessary to meet the

timing requirements while maintaining the required network throughput. The FPGA has the advantage that it can be reprogrammed in the field, simplifying the update and patching process.

4.8.3 Denial of Service Attacks

Denial of service (DOS) attacks are used to prevent users from accessing network resources. A DOS attack is relatively simple to carry out and requires significant resources to counter and recover from such an attack. In a DOS attack, the attacker sends a large number of messages to the machine under attack. Often the attacker uses a large group of machines that they have compromised to conduct the attack and increase the difficulty in stopping the attack. In the case of the substation, the goal of a DOS attack most likely would be to slow the delivery of critical messages between the substation and utility enough to disable the remote control and monitoring capabilities the utility has over the substation.

High-profile websites employ large numbers of web servers to handle their normal load and can switch loads between servers to counter a DOS attack. In this case, those servers under attack will not be used and traffic will be redirected to other servers. Firewalls can be used to ignore messages from the attacking machines but with a large number of machines, the firewall must identify which are involved in the DOS attack. These countermeasures require significant investment and network monitoring. This investment negates some of the benefits of using a commercial network to link the substation to the utility.

One solution to the DOS threat is to use a firewall with strong rules and dedicated IP addresses for the substation and utility. The dedicated IP addresses allow firewall rules to be developed to only accept traffic from a specific set of IP addresses. This will help prevent DOS attacks from getting inside the substation or utility, but will still cause problems at the firewalls where the substation and utility connect to the commercial network. Another solution is a dedicated phone line, cellular link, or satellite link that supports high-speed data transfer, e.g. DSL. This link could be used in emergency situations to provide basic monitoring and control functionality from the utility. The wireless long-range links provide the benefit of being able to function in the event of a disruption of the communication infrastructure. This is especially true for the satellite link. The redundant communication links also provide some protection because an attacker must identify all the links to completely block communication.

4.8.4 Insertion of Malicious IEDs

Substation networks assume that all participants on the network are legitimate IEDs. The presence of a malicious or fake node or a legitimate node that has been compromised provides the attacker with a significant abilities. Through the malicious node, the attacker can inject false information into the substation (Stouffer et al. 2011).

This false information can cause the substation to react to a false set of conditions and provide non-optimal output and in extreme cases even cause a blackout (Dondossola et al. 2009; Stouffer et al. 2011). In the blackout scenario the false information makes the substation believe that a substantial fault is present causing the substation to isolate a sector to protect the rest of the grid. Alternatively, the malicious device can be used to alter the operating condition of the substation to physically damage equipment (Stouffer et al. 2011). This is more problematic because the damaged equipment must be repaired or replaced. The resulting service interruption will take longer and be more costly to correct. Strong authentication methods, such as challenge-response, that meet the timing requirements are needed.

4.8.5 Alteration of Operating Parameters

The operating parameters of the IEDs can be configured remotely using IEC 61850. Of particular interest are parameters relating to the control or operation of the IED, such as set points or an operating mode setting (e.g. normal or test mode flag). Often set points are updated regularly by the control system as it responds to the current electric demand and conditions. Operating mode parameters can be used to set the IED into a diagnostic or test mode during a routine inspection, outage, or during maintenance. If an attacker can access and change these parameters, they can cause blackouts and/or significant damage to equipment (Stouffer et al. 2011). Strong authentication is required for accessing the operating parameters. Some settings such as switching the IED to diagnostic mode could be restricted to manual operation, e.g. a switch, at the substation. Typically, personnel will be present at the substation during maintenance and routine outages to conduct the testing. Hence, the physical security employed at the substation will deter this type of attack. Multiple levels of access requiring additional passwords can help address this threat (Anderson and Leischner 2007).

4.9 Testbeds

Testbeds are important in the evaluation of systems under controlled conditions and without the risk of causing harm to physical equipment. A number of testbeds exist for Smart Grid related applications. These testbeds fall into three categories: real-world testbeds, hardware-in-the-loop testbeds, and simulation testbeds. Real-world testbeds are built using actual components and equipment that are isolated from the larger system. In addition, a testbed environment allows new technologies and security measures to be tested for backward compatibility with legacy devices. Because it is not feasible to upgrade the entire power grid in one iteration, new technologies and security measures must not negatively affect existing devices present in the system. Real-world testbeds provide the best accuracy because real devices are used, but are costly to build and do not easily support investigation of scenario variations. Simu-

lation testbeds are contained entirely in software and use mathematical or discrete event models to model physical and digital processes. Hardware-in-the-loop testbeds are a hybrid of real-world and simulation testbeds, incorporating equipment into a software simulation. Hardware-in-the-loop testbeds simplify evaluations of systems as they scale up in size. The hardware components provide the physical response to the current state of the system and this response is fed into the simulation. The system is scaled up by replicating this behavior within the simulation in software. This type of testbed is useful in analyzing the behavior of a small group of devices or small piece of the larger system.

A simulation testbed for the distribution network for the power grid has been developed by University of Chicago to study the effect of DOS attacks against the Smart Grid communication network on electricity distribution (Davis et al. 2006). This testbed enabled the simulation of a large network where the attacker used the DOS attack to prevent the utility from receiving updated load information showing an increase in load resulting in a blackout (Davis et al. 2006). A defense against the DOS attack based on filtering or blocking the attacking computers was developed and tested yielding promising results allowing the updated load information to reach the operator in time (Davis et al. 2006).

Connecting multiple testbeds together improves the quality and capability of the test by leveraging each testbed's strengths. The Virtual Power System Testbed (VPST) is a hardware-in-the-loop testbed that supports an interface to connect to other testbeds (Bergman et al. 2009).

Another hardware-in-the-loop testbed incorporates an IEC 61850 overcurrent relay with a simulated network and computer workstations (Hahn et al. 2010). This testbed was developed to evaluate a method to search a substation network for NERC CIP (North American Electricity Reliability Council 2009) violations and the proposed method was successful in finding NERC CIP violations (Hahn et al. 2010). NERC CIP is a set of standards providing regulations, best practices, and guidance for security the power grid.

5 Future Work

The use of IEC 61850 to automate and provide remote control to substations requires careful review and reevaluation of operational challenges as well as compliance with existing regulations. Networks must be evaluated to quantify the effect variables such as traffic load, topology, and security requirements have on quality of service and design requirements. This evaluation could be conducted using discrete event simulations and testbeds to verify simulation results. Algorithms for handling missing data items and data received out of order must be researched and developed.

Revision of the regulations and best practices for providing redundant devices will be required as control and protection functions are combined into a single IEC 61850 device. The security implications of this combination must also be researched and the findings reflected in the revised regulations and updated best practices.

Security for the Smart Grid is a critical area of research. Solutions to protect availability of devices, integrity of data, and confidentiality of the data meeting the requirements of the Smart Grid must be researched. Traditional information technology (IT) solutions will not meet the timing requirements of the Smart Grid. Furthermore, solutions with availability of devices as the first priority must be established. This differs from the typical IT case where confidentiality, e.g. for financial transactions, is typically the primary concern. Best practices must be developed to address the security needs while meeting the operational requirements of the Smart Grid.

Standards and best practices for reliability and certification of wireless Smart Grid equipment must be evaluated and revised (Yu and Johnson 2011). Standards from many fields must be analyzed (e.g. electrical protection, EMI, shock, temperature, and humidity) and requirements harmonized for operation of wireless devices in the Smart Grid. These testing and reliability standards must also be converted for use with non-wireless digital systems that will be found within the Smart Grid.

6 Conclusions

The Smart Grid promises to improve electricity delivery, reduce cost, and conserve resources. However, there are many challenges that must be addressed. This chapter has presented the challenges regarding substation automation using the IEC 61850 family of standards. Implementation and deployment of IEC 61850 will require revisions to existing best practices and new best practice guidelines to be developed. These revisions will include topics from the fields of electrical engineering—power engineering and power system protection, and computer science—network architecture and security. A joint effort will be required to address the challenges associated with moving to IEC 61850 and meeting the requirements to provide safe and reliable electricity.

References

Anderson D, Leischner G (2007) Cybersecurity as part of modern substations. Schweitzer Engineering Laboratories, Inc. http://www.selinc.com/WorkArea/DownloadAsset.aspx?id=3530. Accessed 21 Jan 2011

Bergman DC, Jin D, Nicol DM, Yardley T (2009) The virtual power system testbed and inter-testbed integration. In: Proceedings of the 2nd conference on cyber security experimentation and test, USENIX Assoc., Berkeley, CA, USA, 2009

Cortes C, Vapnik K (1995) Support-vector networks. Mach Learn 20(3):273–297

Davis CM, Tate JE, Okhravi H, Grier C, Overbye TJ, Nicol D (2006) SCADA Cyber security testbed development. In: 38th North American Power Symposium, pp 483–488, Carbondale, IL, 17–19 Sept 2006

Deshpande JG, Kim E, Thottan M (2011) Differentiated services QoS in smart grid communication networks. Bell Labs Tech J 16(3):61–81

Dondossola G, Deconinck G, Garrone F, Beitollahi H (2009) Testbeds for assessing critical scenarios in power control systems. In: Setola R, Geretshuber S (eds) Critical information infrastructure security. Lecture Notes in Computer Science, vol 5508. Springer, Heidelberg, pp 223–234

Forouzan BA (2003) TCP/IP Protocol suite, 2nd edn. McGraw-Hill, New York, NY

Gao H, Jin W, Liu G (2008) Simulation study on delay of end-to-end data communication for protective relaying in substations. Front. Electr. Electron. Eng. China 3(2):246–250

Giani A, Karsai G, Roosta T, Shah A, Sinopoli B, Wiley J (2008) A testbed for secure and robust SCADA systems. SIGBED. doi:10(1145/1399583):1399587

Hahn A, Kregel B, Govindarasu M, Fitzpatrick J, Adnan R, Sridhar S, Higdon M (2010) Development of the PowerCyber SCADA security testbed. In: Sheldon FT, Prowell SJ, Abercrombie RK, Krings A (eds) Proceedings of the sixth annual workshop on cyber security and information intelligence research. ACM, New York, NY, USA

Higgins N, Vyatkin V, Nair N-KC, Schwarz K (2011) Distributed power system automation with IEC 61850, IEC 61499, and intelligent control. IEEE Trans Syst Man Cybern Part C Appl Rev 41(1):81–92

Hossenlopp L (2007) Engineering perspectives on IEC 61850. IEEE Power Energ Mag 5(3):45–50

IEEE (2008) IEEE 1588–2008—IEEE standard for a precision clock synchronization protocol for networked measurement and control systems, IEEE.

Ingram M, Ehlers R (2007) Toward effective substation automation. IEEE Power Energ Mag 5(3):67–73

Jung S, Song J, Kim S (2008) Design on SCADA test-bed and security device. Int J Multimedia Ubiquitous Eng 3(4):75–86

Kanabar MG, Sidhu TS (2011) Performance of IEC 61850–9-2 process bus and corrective measure for digital relaying. IEEE Trans Power Delivery 26(2):725–735

Kanabar MG, Sidhu TS, Zadeh MRD (2011) Laboratory investigation of IEC 61850–9-2-based Busbar and distance relaying with corrective measure for sampled value loss/delay. IEEE Trans Power Delivery 26(4):2587–2595

Mackiewicz R (2011) Technical overview and benefits of the IEC 61850 standard for substation automation. http://www.sisconet.com/downloads/IEC61850_Overview_and_Benefits_Paper_General.pdf. Accessed 5 Nov 2011

Men D-Y, Liu W-Y (2011) Application of least squares support vector machine(LS-SVM) based on time series in power system monthly load forecasting. In: 2011 Asia-pacific power and energy engineering conference, Wuhan, China, 25–28 March 2011

Mercurio A, Di Giorgio A, Cioci P (2009) Open-source implementation of monitoring and controlling services for EMS/SCADA systems by means of web services—IEC 61850 and IEC 61970 standards. IEEE Trans Power Delivery 24(3):1148–1153

Moreno N, Flores M, Torres L, Juárez J, González D (2010) Case study: IEC 61850 as automation standard for new substations at CFE, practical experiences. Schweitzer Engineering Laboratories, Inc. http://www.selinc.com/WorkArea/DownloadAsset.aspx?id=7405 Accessed 23 Jan 2012

Myrda P, Donahoe K (2007) The true vision of automation. IEEE Power Energ Mag 5(3):32–44

North American Electricity Reliability Council (2009) NERC Critical infrastructure protection (CIP) Reliability standards, North American Electricity Reliability Council

Ozansoy CR, Zayegh A, Kalam A (2009) The application-view model of the international standard IEC 61850. IEEE Trans Power Delivery 24(3):1132–1139

Rietmann P, Reinhardt P (2006) Applying IEC 61850 to substation automation systems. Paper presented at the PSP, (2006) Power System Protection Conference, Bled. Slovenia. 6–8 Sept 2006

Sanchez G, Gomez I, Luque J, Benjumea J, Rivera O (2010) Using internet protocols to implement IEC 60870–5 telecontrol functions. IEEE Trans Power Delivery 25(1):407–416

Sidhu TS, Yin Y (2007) Modelling and simulation for performance evaluation of IEC61850-based substation communication systems. IEEE Trans Power Delivery 22(3):1482–1489

Stallings W (2000) Data and computer communications, 6th edn. Prentice Hall, Upper Saddle River, NJ

Stouffer K, Falco J, Scarfone K (2011) Guide to industrial control systems (ICS) security. NIST Special Publication 800–82, Gaithersburg, MD

Wang W, Xu Y, Khanna M (2011) A survey on the communication architectures in smart grid. Comput Netw 55:3604–3629

Xiong X, Yu J, Liu X, Shen Z (2008) Reliability of substation protection system based on IEC61850. Trans Tianjin University 14(2):118–122

Yu Q, Johnson RJ (2011) Smart grid communications equipment: EMI, safety, and environmental compliance testing considerations. Bell Labs Tech J 16(3):109–131

Phasor Measurement Unit and Phasor Data Concentrator Cyber Security

Thomas H. Morris, Shengyi Pan, Uttam Adhikari, Nicolas Younan, Roger King and Vahid Madani

Abstract Future bulk electric transmission systems will include substation automation, synchrophasor measurement systems, and automated control algorithms which leverage wide area monitoring systems to better control the grid. Prior to installation of new networked devices, utilities should perform cybersecurity testing and develop corrective actions for identified vulnerabilities. This chapter discusses a set of cyber security requirements developed for phasor measurement units and phasor data concentrators, two intelligent electronic devices which will be added to utility networks to implement synchrophasor systems. Five classes of cyber security requirements are presented; compliance, access, availability, integrity, and confidentiality. Next this chapter discusses testing multiple phasor measurement units and phasor data concentrators from multiple vendors. The test section provides guidance on testing for each requirement and discusses generalized test results.

Keywords Phasor measurement unit · Phasor data concentrator · Cyber security requirements · Cyber security testing

1 Introduction

Many utilities in the United States of America received grants from the Department of Energy under the American Recovery and Reinvestment Act (ARRA) to create wide area monitoring systems. The ARRA grants require recipient utilities to develop a cybersecurity plan which includes a risk assessment as part of parent wide area monitoring systems projects. Wide area monitoring systems require installation of

T. H. Morris (✉) · S. Pan · U. Adhikari · N. Younan · R. King
Mississippi State University, Mississippi, MS, USA
e-mail: morris@ece.msstate.edu

V. Madani
Pacific Gas and Electric Company, San Francisco, CA, USA

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_7,

phasor measurement units (PMU), and substation phasor data concentrators (PDC), among other devices and software. PMUs and substation PDCs are networked appliances which use routable protocols. As such, these devices may be declared North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection (CIP) standard 002-3 critical cyber assets (CCA), depending upon each individual unit's application within the power system. CCA must be housed within an electronic security perimeter and undergo a cyber vulnerability assessment.

PMU and substation PDC are networked appliances and may become the target of attacks against bulk electric power systems. Threats against these devices include denial of service attacks, attacks against open ports and services intended to elevate privilege, attempts to change device settings, attempts to inject malicious device commands, attempts to hijack device access credentials or other confidential information, and attempts to place a man-in-the-middle between devices.

Section 2 of this document provides a brief review of related works. Section 3 describes a set of cyber security requirements for PMU and PDC. Three primary sources were used to derive cyber security requirements; NISTIR 7628: Guidelines for Smart Grid Cyber Security, Department of Homeland Security: Cyber Security Procurement Language for Control Systems, and a set of internal requirements from the utility. Five classes of cyber security requirements are offered; compliance, access, availability, integrity, and confidentiality. For each class individual requirements are provided and discussed. After introducing the cyber security requirements Sect. 4 describes a test process to confirm PMU and PDC meet the listed requirements. The testing section is organized by class. Generalized test results, with guidance on what utilities should expect when testing, are provided for the individual requirements. Some results are not discussed in detail to avoid breaching confidentiality agreements and to avoid enabling cyber attacks. The section discussing availability provides and extended overview of network congestion testing, denial of service testing, and protocol mutation testing performed using a Spirent (formerly MU Dynamics) MU-4000. Finally, the chapter provides a section of conclusions.

2 Related Works

The Idaho National Labs (INL) National SCADA Testbed Program is a large scale test bed program dedicated to control system cybersecurity assessment, standards improvement, outreach, and training. Noted research outcomes from the INL SCADA Testbed Program include published taxonomies of common industrial control vulnerabilities, published lessons learned from security assessments control systems (Fink et al. 2006), participation in standards enhancement and development, and development of recommended procurement language for wireless systems in the advanced metering infrastructure (2009). INL activities primarily involve security assessments, outreach, training, and standards development for the electric power industry. INL partners with industry software and equipment vendors for cyber security assessments of products. This chapter discusses a alternative method for developing PMU

and PDC cyber security requirements and testing to determine requirement compliance.

Researchers have performed vulnerability assessments of generation and substation devices to support development of taxonomies of vulnerabilities related to industrial control systems. In (Fovino et al. 2010) Fovino et al. describe a test bed used for vulnerability assessment of components found in a Turbo-Gas Power Plant. In (Skaggs et al. 2002), Skaggs et al. describe a tool, NETGLEAN, testing device for network vulnerabilities. Two well known tools are available for network vulnerability testing of industrial control systems. These vulnerability assessments are used by committees such as the cyber security working group which wrote the NISTIR 7628 Guidelines for Smart Grid Cyber Security and the NERC CIP drafting team which write the NERC CIP standards. Both the NISTIR 7628 guidelines and the NERC CIP standards were used in this work to develop a set of cyber security requirements and recommendations for PMU and PDC.

Wurldtech (Wurldtech Security Technologies) offers the Achilles Satellite product for testing industrial control system devices. Spirent (formerly MU Dynamics) offers the Spirent Studio Test Suite for testing networked devices, include industrial control system devices. Both products include protocol mutation and denial of service test suites. A MU-4000 Network Analyzer was used in testing described in this chapter to perform denial of service testing. Tests described in this work which leveraged the MU-4000 include network congestion testing, protocol mutation testing, known vulnerability exploit testing.

3 Synchrophasor System Cyber Security Requirements Development

A set of cybersecurity requirements and recommendations were prepared from review of the National Institute of Standards and Technology Interagency Report (NISTIR) 7628 Guidelines for Smart Grid Cyber Security: Vol. 2, Security Architecture and Security Requirements, Department of Homeland Security (DHS): Cyber Security Procurement Language for Control Systems, and utility internal requirements. NISTIR 7628 Vol. 2 includes a process for deriving cyber security recommendations and requirements for smart grid systems. NISTIR 7628 requirements and recommendations are taken from NIST SP 800-53 Revision 3, the Department of Homeland Security Catalog of Control Systems Security: Recommendations for Standards Developers (2009), NERC CIP (1–9). Each requirement is traceable to one or more of the aforementioned source documents.

PMU and PDC will likely be considered cyber critical assets due to their use in Wide Area Monitoring Systems (WAMS), Wide Area Protection Systems (WAPS), and System Integrity Protection Schemes (SIPS). Classification as a cyber critical asset impacts infers a set of 5 classes of cyber security requirements; compliance, access enforcement, availability, control and measurement integrity, and confiden-

Table 1 Cyber security requirement classes

Requirement class	Requirements	
Compliance	Define auditable events	Audit record retention
	Audit record contents	List required ports and services
	Time stamp	Patch management
Access	Enforce access authorizations	Least privilege
	Unsuccessful login attempts	Previous logon notification
	Session lock	Password complexity
	Boundary protection	Intrusion Detection/Monitoring
	Disable unneeded ports and services	Appropriate use banners
	Vulnerability assessment	
Availability	Recovery and reconstitution	Denial-of-service protection
	Audit log performance impact	
Integrity	Non-repudiation	Protection of audit information
	Device identification and authentication	Communication integrity
	Trusted path	Message authenticity
Confidentiality	Communication confidentiality	Use of validated cryptography
	Confidentiality of information at rest	

tiality. Table 1 list the cyber security requirement classes and for each class provides a list of individual requirement names. The set of requirements described here are not exhaustive. This document is restricted to cyber security requirements relevant to phasor measurement units and phasor data concentrators. Requirements related to corporate policy, training, physical security are not addressed except where the PMU or PDC may impact the utilities ability to meet that requirement.

PMU and PDC should ensure cyber security events are appropriately documented to enable effective forensic analysis, to ensure appropriate corrective actions are taken after an incident, and to prepare for audits. Utilities should work with device vendors to *define auditable events*. The set of auditable events will depend upon the device functionality. Common auditable events include events to record device access, events to note device settings changes, events to note security parameter changes, and events to note security related incidents such as password failures and device lock out. A*udit record contents* should include the date and time of the event, the component affected, a descriptive name of the event, a user name if available, and information about the event outcome. The logged event *time stamp* should be based off a system clock so that an event time line can be developed if multiple devices are affected. NERC CIP requires utilities to store audit records for varying amounts of time. The minimum storage times are typically 90 days. Information related to cyber incidents must be stored for 3 calendar years. The PMU and PDC may or may not store audit records in a manner which complies with NERC. The PMU and PDC must at least store audit information temporarily and provide a means to transfer logs to a historian. For audit compliance utilities must prepare a list of used ports and services and note whether ports and services are used for normal operation, emergency operation, or both. PMU and PDC will provide ports and services for

multiple use cases. Utilities should determine which ports and services are required for normal and emergency operation and only list these. Utilities must also have a *patch management* plan. This requirement ensures that utilities review patches from device vendors, test patches before installation, and install patches in a timely manner where technically feasible.

PMU and PDC should limit access to authorized users and systems. The concept of *least privilege* should be enforced. Devices should assign and enforce the most restrictive set of rights and privileges or access needed by users for the performance of specified tasks. Many devices use passwords to *enforce access authorizations*. When passwords are used a minimum *password complexity* standard should be created and enforced. NERC CIP requires a 6 character password with a combination of alpha, numeric, and "special" characters. Password complexity requirements are based upon expected adversaries' ability to reverse engineer the password using brute force methods. Password standards therefore should evolve over time to stay ahead of adversary capabilities. *Unsuccessful login attempts* should result in a logged in event and the log should be reviewed regularly.

Multiple *Unsuccessful login attempts* should result in the user being locked out of the device temporarily. The number of unsuccessful logon attempts which results in a locked system and the duration of the system lock out should be defined by the organization. Care should be taken not to create a denial of service by locking the device for too long a period. When a user successfully logs on to a device an *appropriate use banner* should be displayed and a *previous logon notification* should be displayed. The appropriate used banner warns users against inappropriate activity and describes appropriate uses for the device or software. The previous logon notification provides a time stamp of the last logon and displays the number of unsuccessful logon attempts immediately prior to the current successful logon.

NERC CIP requires cyber critical assets to be placed in an electronic security perimeter. The electronic security perimeter provides *boundary protection*. PMU and PDC provide access control and security event logging capabilities to meet the electronic security perimeter requirement. PMU and PDC will be placed within the electronic security perimeter which generally means a firewall will isolate the PMU and PDC from outside networks. Unused ports and services on PMU and PDC should be disabled. It is preferable to disable unused ports and services within the PMU or PDC themselves since once disabled no device or user will be able to access these ports and services. However, if this is not feasible to disable these at the device, unused ports and services may be blocked by the firewall. However this approach will not prevent devices within the electronic security perimeter from accessing enabled unused ports and services. Electronic security perimeter access points should be monitored for electronic intrusion attempts. PMU and PDC should provide security log information on auditable events to enable this required monitoring. Utilities must subject cyber critical assets to vulnerability assessments. The *vulnerability assessment* requires utilities to discover all access points into the electronic security perimeter. This is generally done by performing port scans from outside the electronic security perimeter to identify open ports and services. Only required ports and services should be found during a port scan.

The availability cyber security requirements ensure the device remains available to perform its primary function. PMU and PDC are networked devices which continually stream power system measurements (voltage and current phasors) to upstream devices for power system monitoring and control. PMU and PDC network services must remain available in the face of *denial of service* attacks. Additionally, the *audit log performance impact* requirement ensures device PMU and PDC device primary functionality is not negatively impacted by logging requirements. Finally, *recovery and reconstitution* ensures that a plan exists to restore PMU and PDC after a cyber attack, power loss, or system reset.

The integrity cyber security requirements ensure that network traffic sent to and from PMU and PDC is unaltered, validate the identification of communicating parties and devices, and ensure logged records are unaltered. The *non-repudiation* requirement ensures that audit records cannot be altered or erased. When feasible a trusted path should be established between the user and the PMU and PDC. PMU and PDC should use digital signatures to provide communication integrity. Synchrophasor measurements may be transmitted wholly within a utility network or may be transmitted to super PDC which collects and disseminates synchrophasor data within an interconnection. *Communication integrity* and *message authenticity* can both be achieved by cryptographically signing IEEE C37.118 data frames. Commands sent to PMU and PDC, using IEEE C37.1118 and all other ports and services supported by the PMU or PDC, should also be cryptographically signed to ensure Communication integrity and message authenticity. Machine to machine connections should also use digital signature technologies to ensure communication integrity and message authenticity between machines.

The confidentiality cyber security requirements are limited for PMU and PDC. Generally utilities do not consider PMU and PDC synchrophasor measurements, commands, and settings confidential information. One exception is confidentiality of user credentials. These confidentiality requirements extend to network transmission of passwords for remote access and safe storage of passwords within a device. Passwords should not be transmitted over network interfaces or stored directly on a device. Symmetric or asymmetric ciphers should also not be used to protect password confidentiality. A one-way function such as a cryptographic hash function should be used to encrypt the password before transmission or storage. Passwords should also be salted before applying the hash function. Older devices also commonly store and transmit passwords as plain text or use non-validated cryptographic methods which only obfuscate the passwords for storage. A common scheme to protect passwords in many industrial control system devices is to XOR a password with a fixed number. This scheme is easily reversible.

Both the integrity and confidentiality requirements rely on the use of cryptographic functions. PMU and PDC should only use validated cryptographic functions. A common requirement for validated cryptographic functions is to limit cryptographic function to those approved by NIST Federal Information Processing Standards (FIPS).

4 Synchrophasor System Cyber Security Component Testing

The cyber security requirements from the above section were applied to hardware, software, and communication systems throughout the synchrophasor system. A diagram was developed which included all system components and communication interfaces to each component. A sanitized version of the synchrophasor system component diagram is shown in Fig. 1. Cybersecurity requirements conformance was handled differently for different sub-systems. Energy management system (EMS) conformance and testing was assigned to the EMS software vendor. Phasor measurement unit (PMU) and Phasor data concentrator conformance and testing was performed in two steps. First, the PMU and PDC hardware vendors performed cyber security testing in house. Second, third party cyber security testing was performed on the PMUs and PDCs. Cyber security test reports were submitted to the utility and PMU and PDC vendors for review. Cyber security test reports included test results with vulnerabilities ranked using a risk scale proprietary to the utility. All vulnerabilities were addressed by the cyber security team by either changes to firmware executed on the PMU and PDC or by system level architecture changes.

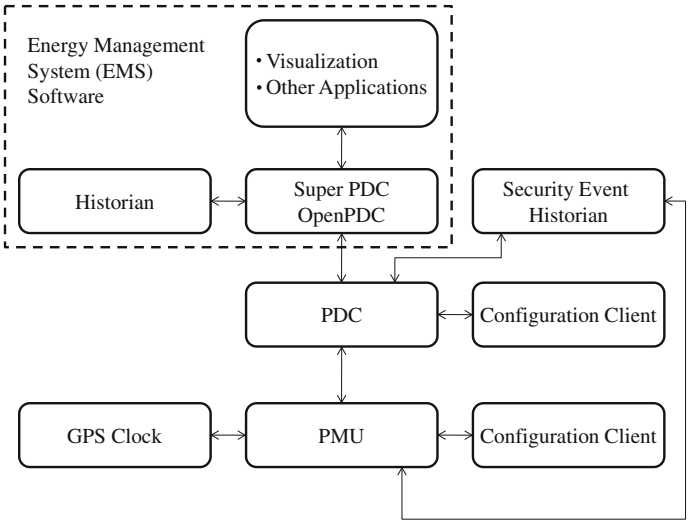

Fig. 1 Synchrophasor system component diagram

4.1 Compliance

PMU and PDC user manuals were reviewed for cybersecurity features. Identified cyber security features were listed in the PMU and PDC test reports. Cyber security features were also tested in a laboratory setting to confirm functionality.

Define auditable events-To aid the utility in defining auditable events the PMU and PDC user manuals were reviewed. All security related events were identified. For phasor measurement units and phasor data concentrators, auditable cyber security events are typically limited to invalid access attempts. These devices typically require password entry or another form of authorization before a user can perform restricted actions. Restricted actions include changing device settings, sending the device commands, changing security parameters, and altering device logs. When a user attempts to take a restricted action and after authorization failure (eg. password failure) the invalid access attempt is logged. The *audit record contents* and *time stamp* requirements define what information is logged. Logged information must at least include the day and time of the event, the type of event that occurred, and information on which user performed the invalid action.

Audit record retention-NERC CIP requires utilities to store audit records for varying amounts of time. The minimum storage times are typically 90 days. Information related to cyber incidents must be stored for 3 calendar years. The tested PMU and PDC were not capable of storing audit records for these durations. The tested devices used event logs which contained security and non-security audit information. The logs were designed with a first in first out (FIFO) storage mechanism in which older log entries were erased when the log over flowed. Testers should be aware of this type storage architecture and recommend a means of transferring security records to a historian designed to meet NERC CIP requirements.

Audit record contents-The tested PMU and PDC logged security events by name and included a time stamp. The tested PMU are designed to be accessed locally via a faceplate and remotely via a manufacturer provided software client. Audit record contents differ according to whether the device was accessed locally or remotely. Local access does not require a user ID for sign on. Therefore local records do not include user information related to security events. Remote access requires sign on with user ID to the remote software client. As such audit records for remote access include user information where relevant.

Time stamp-The tested PMU and PDC included time stamps for each audit record. The timestamp was derived from the device's system time. Many substation devices include mechanisms to set time and date by GPS clock or by network services. PMU and PDC clocks are set within 1 microsecond accuracy of Universal Time Coordinate. PMU and PDC should use the same clock source for audit logs and for time stamping synchrophasor measurements.

List required ports and services- Utilities should work with PMU and PDC providers to develop a list of required ports and services. Tools such as NMAP and OpenVAS can be used to identify ports and services for devices. However, these tools will list all ports and services regardless of whether they are required or not.

Often hardware providers will include ports and services for multiple possible use cases. For example, a PMU or PDC may include TELNET, MODBUS, and DNP3 to provide remote access. A utility will likely only need one means for remote access. The others should be disabled.

Patch management-Patch installation for field devices is time and resource intensive. Also phasor measurement units are integrated into existing protection relays. Utilities are typically hesitant to patch protection relays except for the most serious security vulnerabilities.

4.2 Access

Enforce access authorizations-Access enforcement for field devices is commonly broken into local and remote access. Local access is access from the faceplate of a device. Remote access is access over a computer network or from a computer directly connected to the device. Remote access and local access via a connected computer typically occurs through the use of an associated software client. Devices should implement least privilege access restrictions. PMU and PDC device behavior varies according to whether access is local or remote. Devices typically compartmentalize interaction into groups. Interaction groups may include settings changes, commands, and administrative. For PMU and PDC settings actions include changing PMU or PDC behavior such as the number of samples per second, the number and types of measurements, interpolation settings, and missing data settings. Settings also include security related items such as changing passwords, changing the minimum number password failures to trigger a lock out, and changing the lockout duration. Commands include items such as clearing event logs, deleting individual records, and forcing a new PMU measurement. Local users will have to enter a password before performing privileged operations in the settings, command, and administrative categories. Three schemes for remote access control were seen during testing. First, some devices required password entry to raise privilege level similar to the local access case. Second, some devices required username and password and utilized a role based access control scheme for remote access. Thirds, some devices used the Remote Authentication Dial In User Service (RADIUS) to authenticate users before allowing access. A RADIUS or similar scheme is preferred since ID verification is more robust in such a scheme.

Least privilege-All PMU and PDC tested met the principle of least privilege by including multiple levels of privilege. Devices defaulted to the lowest privilege level. For local access users were automatically logged out after a utility defined time period. For remote access users typically stayed logged in indefinitely. Utilities should require that remote users are also logged out after a utility defined period.

Password complexity-NERC CIP requires a 6 character password with a combination of alpha, numeric, and "special" characters. Test results varied by device. Some devices did not meet the NERC CIP minimum for either local or remote passwords or both cases. Some devices would allow a password which met the standard, but,

did not require it. Utilities should work with device vendors to ensure both local and remote passwords meet NERC CIP requirements. Utilities may receive an exception allowing non-compliant passwords for older devices, however, over time newer devices should come into compliance.

Unsuccessful login attempts-PMU and PDC should automatically lock out access for a utility defined period of time after a utility defined number of incorrect login attempts. All tested devices supported this feature. Utilities should set the user lock out to 3–5 min. Avoid longer times as this could create a denial of service. Additionally, there should be a mechanism to override the lock out period during an emergency.

Session lock-PMU and PDC should log out a user after a utility defined period of inactivity. Tested PMU and PDC had mixed results with this feature. Some devices log out local users after a period of inactivity but did not log out remote users. Utilities should work with device vendors to ensure that both local and remote users are automatically signed off after a period inactivity.

Appropriate use banners-PMU and PDC should display an appropriate use banner when logging in. Tested devices supported displaying an appropriate use banner when logging in via remote connection. PMU and PDC faceplates are not conducive to displaying a banner before local access. Utilities should work with device vendors to ensure that the appropriate use banner is customizable.

Boundary protection-PMU and PDC used in wide area monitoring (WAMS), wide area protection systems (WAPS), and system integrity protection schemes (SIPS) will be cyber critical assets and therefore be placed within an electronic security perimeter. Some authors (Stewart et al. 2010) have proposed using an IPSEC or similar tunnel to provide authentication and confidentiality services to PDC communications. Tested PDC did not support IPSEC, SSH, or SSL tunnels directly. However, many industrial firewalls are capable of tunneling data from the firewall perimeter to a tunnel enabled device.

Disable unneeded ports and services-NMAP and OpenVAS (NMAP Security Scanner) were used to perform port scans of the tested PMU and PDC. This testing was performed to identify unused ports and services which were not disabled. Utilities will likely need to work with the device provider to disable unused ports and services. A configuration setting may be available at installation time or a special firmware may be required from the provider to disable unused ports and services. If a port cannot be disabled at the device a firewall rule can be developed to deny all access to unused ports. This approach should be considered an exception and avoided unless no other means exists to disable the port. PMU use the IEEE C37.118 network protocol. This protocol can be implemented as a UDP or TCP service. PMU are servers and will have one TCP (port 4712) or UDP (port 4713) port open for the IEEE C37.118 server daemon. PDC are clients connected to multiple PMU. PDC will open TCP or UDP session as a client for each connected PMU. The client ports will not appear on a port scan report since these port numbers will be assigned by the operating system at the time of opening the session. PDC may also act as a server since a PDC may appear as a PMU to upstream PDC. As such a PDC may have one TCP (port 4712) or UDP (port 4713) port open for the IEEE C37.118 server daemon. PMU and PDC

will also include a service for remote control. Common services used for remote control include MODBUS, DNP3, and TELNET. PMU and PDC may implement other ports and services. Each additional port or service must be reviewed to ensure it is necessary for ordinary or emergency operation.

Intrusion Detection/Monitoring-Utilities must monitor all access ports for signs of unauthorized access attempts. At a minimum, utilities should monitor the security alerts from PMU and PDC to detect login failures. Utilities should also use intrusion detection products unauthorized access. Utilities should monitor all PMU and PDC communication traffic. Monitoring can be performed by intrusion detection systems within the firewall, by separate computers in the substation running intrusion detection products such as Snort. Signature based monitoring can include signatures to inspect traffic for known exploits, signatures to detect for exploits against vulnerabilities discovered in product testing, and proactive signatures designed to detect attempted protocol mutation attacks. Utilities should also consider model based and or anomaly based intrusion detection systems to attempt to detect zero day exploits.

4.3 Availability

Recovery and reconstitution-PMU and PDC should recover from power cycling events and loss and restoration of communication network connections. All tested PMU and PDC were subjected to power cycling events. PDC are pre-configured with information about each connected PMU. After power up PDC poll connected PMU and request a configuration frame. After the configuration frame is sent PMU immediately begin transmitting data frames to the IP address of the PDC which requested the configuration frame. All tested PMU and PDC successfully automatically recovered after power events. PMU and PDC were also subjected to multiple types of network connectivity events. First, the cable connecting tested devices to a network switch was disconnected. In the second case, tested devices were placed in a Virtual local area network (VLAN) after power up and communication had commenced. In the third case devices were subjected to a man-in-the-middle attack using Ettercap. After each case devices were reconnected to the network by undoing the mechanism used to disconnect the device. In each case, for each device, the device was able to resume communication after the network connection was reestablished.

Audit log performance impact-Testers provided stimulus to each tested PMU and PDC to exercise all security related audit events. All stimuli were performed while the PMU and PDC were in operation, measuring and streaming synchrophasor packets. No ill effects were noticed in device operation due to security event logging.

Denial of service protection-NISTIR 7628 volume 1 recommendation SC-5 Denial of Service Protection states “The Smart Grid information system mitigates or limits the effects of denial-of-service attacks based on an organization-defined list of denial-of-service attacks” (The smart grid interoperability panel—cyber security working group). This recommendation leaves the process of identifying denial-of-service vulnerabilities to the utility. Two methods were used to identify denial of

service vulnerabilities related to the installation of PMU and PDC. First, network congestion tests were performed to test the device and system's ability to handle high volumes of network traffic. The network congestion tests also include well known denial of service exploits (such Ping flood, Teardrop, LAND attack, etc). Second, protocol mutation testing was performed to attempt to identify unknown denial of service vulnerabilities specific to the tested PMU and PDC. Sects. 4.3.1–4.3.3 detail the PMU and PDC denial of service testing.

4.3.1 Test Configuration

A Spirent (formerly MU Dynamics) MU-4000 Analyzer was used to perform denial of service, network congestion, and protocol mutation tests. A personal computer (PC) was used with Wireshark to capture network traffic data logs and to host software used to configure and remotely control the PMU and PDC. The PMU were connected to a Real Time Digital Simulator (RTDS) in a hardware-in-the-loop configuration. The RTDS provided simulated high voltage AC busses for the PMU's to measure. PMU were connected through a substation router to PDC. PDC concentrated synchrophasor measurement streams from the PMU and forwarded this data to an OpenPDC installation which served as a historian for the system. Figure 2 shows the test bed configuration.

PMU periodically (typically at 30, 60, or 120 Hz) measure voltage, current, and transmit voltage and current phasors (based upon a reference cosine waveform). PMU are time synchronized devices with clocks synchronized to Universal Time Coordinated (UTC) with 1 microsecond accuracy. Synchrophasor network packets are transmitted from the PMU to a PDC. PMU adhere to the IEEE C37.118 standard which specifies measurement requirements and the synchrophasor measurement format. PMU may communicate over Ethernet or Serial port. Three PMU's were tested for this work. PMU A and PMU B shared the same vendor, while PMU C was manu-

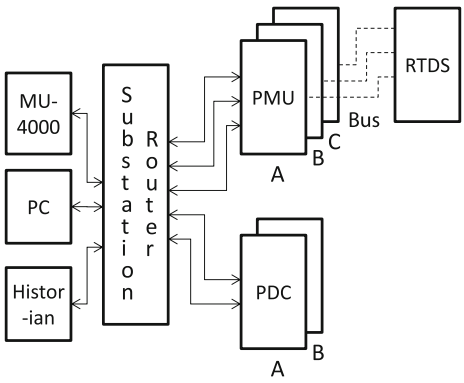

Fig. 2 Test bed configuration

factured by a second vendor. Both PMU communicate over Ethernet using the IEEE C37.118 protocol.

PDC collect synchrophasor streams from multiple PMU and create a single stream for retransmission to another PDC or historian. PDC perform stream data rate conversion and can be configured to interpolate when data is missing from a stream. PDC adhere to the IEEE C37.118 standard and communicate over Ethernet. Two PDC were tested for this work. PDC A and PDC B were manufactured by separate vendors.

4.3.2 Network Congestion Testing

The MU-4000 Network Analyzer was used to perform network congestion testing. The MU-4000 denial of service test suite includes tests for multiple network protocols across all network OSI layers. The denial of service tests validate a device's ability to withstand large volumes of traffic directed at the device. The test engineer should identify relevant network protocols for testing.

Each network congestion test attempts to stress a separate portion of the device's network stack. The tests target a device's ability to process large volumes of a single type of network traffic. Many substation network appliances contain limited memory which can be exhausted and lead to operating system exceptions, cause services to stall, and or cause the device to reset itself. A set of network layer tests send floods of ARP requests, PPPOE packets, and IPv4 packets to the target device. Network layer variations send random packets of all three types, IP packets with random sizes and random payload, and IP packets with large numbers of IP fragments. A set of ICMP tests were also used. ICMP tests send floods of ICMP echo requests (aka. Ping flood or Smurf attack), ICMP echo packets with large payloads, address mask requests, and source quench messages.

Transport layer tests send floods of TCP and UDP packets to the device under test. TCP tests include variations which stress a device's ability to create and teardown TCP sessions with floods of TCP SYN and TCP FIN packets targeting individual TCP ports and to random TCP ports. UDP tests include random headers, port numbers, and payloads.

Two tests validate device behavior for illegal packet types. A LAND test sends floods of IP packets with both the source and destination IP address set to the target's IP address. A teardrop test sends fragmented IP packets which have overlapping IP fragments.

All devices tested eventually became unresponsive when the traffic volume increases beyond that devices ability to process packets. Figure 3 shows typical device behavior to denial of service tests. The brown triangle shows the rate packets are being transmitted to the target device. As the tester ramps the packet rate it periodically sends the target an instrumentation packet (a query which the tested device is known to support) to test if the device is able to respond. The instrumentation packet may be a TCP session request on a supported port or an ICMP echo request or any other type of packet the target is known to be capable of responding to. The blue vertical

Fig. 3 Denial of service test response time chart

lines show the target device responding to instrumentation requests. A taller blue line indicates a slower response time. The red dots indicate failed instrumentation request. As the packet rate increases devices become unresponsive. Some devices may hang or reset themselves when subjected to high packet rates. Many devices are unresponsive during the test, but, become responsive again when the packet rate returns to acceptable levels.

Understanding the packet rate which causes a device to become unresponsive is important for system planning and for creating an effective denial of service mitigation approach. Figure 4 shows an example availability chart for a single denial of service test against a device. The availability shows the percent availability (Y-axis), percentage of time that a device is able to respond to instrumentation requests, versus packet transmission rate (X-axis).

Utility engineers and network administrators can use the availability chart to define a maximum threshold for traffic congestion at the switch or router within the substation for the different traffic types. Based upon testing results it is recommended that utilities monitor network traffic volume in control system networks to detect transmission of high volumes of traffic. Monitoring systems should alert a human administrator to enable mitigation. Routers in the control system network may be configured to limit traffic sent to the PMU or PDC or may be configured to close ports sourcing offensive amounts of network traffic. Automatically closing router ports is potentially dangerous since critical traffic may use the port. A thorough system review should be performed before enabling automatic port closure. Maximum traffic rate thresholds should be defined for all relevant traffic types.

PMU and PDC transmit continuous streams of measurements at 30, 60, or 120 samples per second. Measurements are time stamped with 1 microsecond accuracy relative to universal time coordinated (UTC) time. It is important to understand PMU and PDC behavior after DOS event completes. Testers should confirm that tested devices and network appliances in the route do not queue large volume of IEEE C37.118 data packets which then leads to a synchrophasor stream which is perpetually delayed. PDC hold data from on time PMU to wait for data packets from late arriving PMU streams. A denial of service attack can have a persistent effect if the attacked PMU's date stream becomes consistently late after the attack. PDC eventually drop old data packets and begin to interpolate. PMU and PDC which

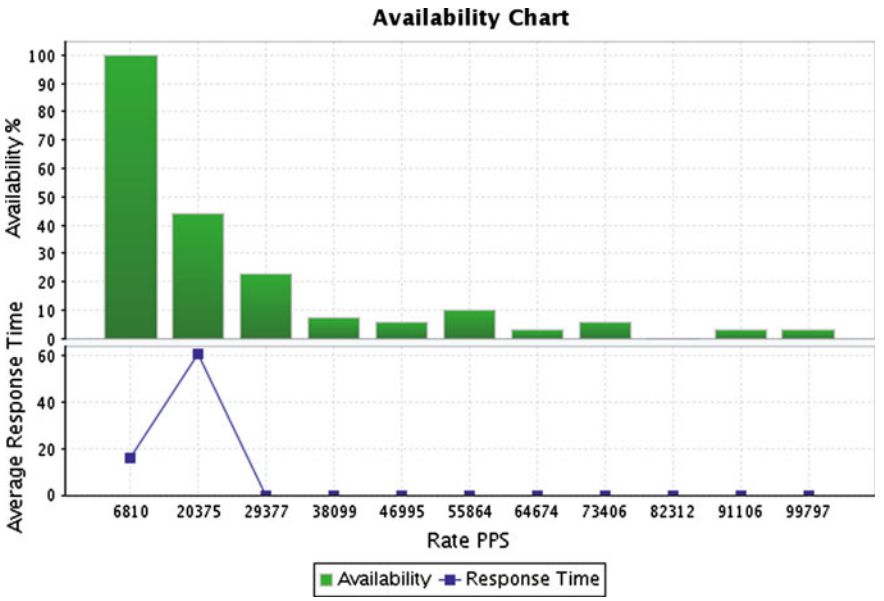

Fig. 4 Availability chart from congestion testing

recover from a denial of service attack should clear their transmit queues to avoid the aforementioned effects.

4.3.3 Protocol Mutation

A second method to test for denial of service vulnerabilities is through protocol mutation, also known as fuzzing. Protocol mutation creates network packets with random contents. Each field in a packet's header, payload, and trailer is assigned a set of variant values. Variant values for a field may include legal values and illegal values. The protocol mutation tester creates a set of packets which include all combinations of all fields with all variant values. The number of combinations grows quickly and protocol mutation can be a slow process. The benefit of protocol mutation is that combinations of fields which may not be thought of by a human can be tested to confirm that the device network stack does not hang or reset when the test packet is processed. Protocol mutation is intended to discover vulnerabilities before they are discovered by an adversary and become exploited zero day vulnerabilities.

The selection of protocols for mutation testing was based on port scanning and device manual review results. All communication protocol supported by a device should be tested. Mutated protocols for the PMU and PDCs included ARP, TCP, UDP, IP, ICMP, DNP3, MODBUS, IEEE C37.118, and HTTP.

The MU-4000 Network Analyzer was used to perform protocol mutation testing. As with the denial of service testing the tester sends groups of mutated packets to the target device. The tester periodically sends instrumentation packets (queries which the tested device is known to support) to confirm that the device under test can still respond. Protocol mutation requires two types of instrumentation packets. The first instrumentation is a communication packet and response pair which is known to work on the target device. This instrumentation is typically unrelated to the mutated protocol. This instrumentation confirms the device network stack is still functioning and responsive. It is possible the portion of the network stack associated with the mutated protocol will hang without affecting other parts of the network stack. For example, a UDP mutation may hang the UDP stack, but leave the TCP stack functioning correctly. The second instrumentation request type is a known good packet of the type being mutated. This instrumentation confirms the portion of the network stack related to the mutated protocol is still functioning and responsive.

Some services were capable of assignment to a variable TCP or UDP port number. In this case, protocol mutation was repeated for multiple ports. A good strategy for testing services with variable ports is to repeat testing with port assigned to multiple port numbers in the well known space (0–1,023), multiple port numbers in the registered port range (1,023–49,151), and multiple port numbers in the private range (49,152–65,535). Some services are capable of assignment to a fixed set of port numbers. In this case, it is good practice to test at all legal port assignments.

The MU-4000 includes built-in protocol mutation capabilities for many well known protocols. Some protocols are not supported. For example, IEEE C37.118 is not natively supported. Also, newly developed protocols may not initially be supported. The MU-4000 is capable of learning protocols from Wireshark packet captures. After learning a protocol the MU-4000 scenario builder can generate protocol mutations to test a device. The scenario builder feature was used for IEEE C37.118 protocol mutation. Only frames received as input by the target device should be mutated and sent to the target device. Mutated IEEE C37.118 commands frames were mutated and sent to PMUs. Mutated IEEE C37.118 configuration and data frames were sent to the PDCs.

Protocol Mutation testing may indentify individual packets which cause device failures including hanging network stacks or causing the device under test to reset itself. Protocol Mutation testing may also indentify combinations of packets which cause similar device failures. In both cases careful study is required to determine the root cause of the failure. Mitigation of detected vulnerabilities can be achieved with a firewall or signature based intrusion prevention system (IPS) rules to block problem traffic. Vulnerabilities identified using protocol mutation should also be reported to the device vendor. Protocol mutation identified multiple issues on devices tested for this work. Issues included crashing of individual network services, crashing of applications running on devices, and unintended soft resetting of affected devices.

4.4 Integrity

Protection of audit information-PMU and PDC should protect audit information from unauthorized access, modification, and deletion. Tested PMU and PDC required privilege escalation, by password entry or other means, before an audit record could be deleted. Audit information may be transmitted from PMU or PDC to security historians using protocols which do not protect message authenticity. In these cases an audit record may be modified during transmission.

Non-repudiation-Users should not be able to alter or erase audit records. PMU and PDC may not comply with this requirement. First, PMU and PDC are resource constrained devices which often do not include facilities to use cryptographic digital signature algorithms to sign audit records. This lack of a cryptographic signature could enable alteration of an audit record. However, this may not feasible for contemporary PMU and PDC. This feature should be added for future devices. Second, PMU and PDC may provide a mechanism to easily erase audit records. This feature, if it exists, should be disabled. Finally, as mentioned above PMU and PDC may have limited audit record storage. Utilities should develop an approach to transfer audit records to a properly configured historian.

Trusted path-Most PMU and PDC do not provide a mechanism to establish a trusted path between the PMU or PDC and the user during remote access. Newer devices support using RADIUS to authenticate users before allowing access to the PMU or PDC. Software clients used to facilitate remote access should also authenticate the identification of PMU and PDC before establishing a connection.

Communication integrity, message authenticity, and *device identification and authentication* features are not natively supported in tested PMU and PDC. PMU and PDC can be configured remotely. Settings and configuration changes were found to be transmitted using common communication protocols such as TELNET, MODBUS, and DNP3. These protocols do not include features for communication integrity and message authenticity. Use of SSL, SSH, or IPSEC will provide communication integrity and message authenticity for older protocols which do not support it natively. PMU and PDC transmit synchrophasor measurements using the IEEE C37.118. IEEE C37.118 also does not include message authenticity features. In 2010 Stewart et al. discuss the feasibility of using IPSEC to protect synchrophasor communications. Feasibility depends upon the applications which will use the synchrophasor measurements. Wide area visualization applications can accept the delay associated with IPSEC. Some wide area protection systems which will use synchrophasor measurements will not be able to accept IPSEC delays. A new standard is in development to add message authenticity and confidentiality features to IEEE C37.118 network traffic. IEC 61850 90-5 packets will encapsulate IEEE C37.118 packets to provide the additional security features.

4.5 Confidentiality

Communication confidentiality- Tested devices did not support encryption to provide PMU and PDC communication confidentiality. A common opinion among industry engineers is that confidentiality is not required in control system networking. At a minimum, passwords should be encrypted when transmitted to PMU or PDC. PMU and PDC tested were found not to encrypt passwords during transmission. Passwords were transmitted as plain text or obfuscated with XOR schemes not based on approved cryptographic methods.

Confidentiality of information at rest-Many PMU and PDC store passwords using two way algorithms. This is done to enable password recovery. Passwords should not be stored directly or after encryption with a two way algorithm. Passwords should be salted and hashed before storage on a PMU or PDC.

Use if validated cryptography- Use of cryptographic algorithms is rare for PMU and PDC. As devices add communication integrity, trusted path, and confidentiality features, PMU and PDC should use FIPS approved cryptographic algorithms.

5 Conclusions

Synchrophasor systems are an emerging technology. Prior to installation of a synchrophasor system a set of cyber security requirements must be developed, new devices must undergo vulnerability testing, and proper security controls must be designed to protect the synchrophasor system from unauthorized access.

In this chapter we described the process used to develop a set of cyber security requirements in the design stage of a synchrophasor project. Five classes cyber security requirements are provided in this chapter; compliance, access, availability, integrity, and confidentiality. For each requirement class separate requirements are provided. In total 29 requirements are discussed from the 5 requirement classes. Each requirement is explained in detail. Next, the chapter discusses a cyber security vulnerability analysis and testing process. This section is organized to provide a description of test processes and generalized results for each cyber security requirement. Extra detail is provided network congestion and protocol mutation testing of multiple phasor measurement units and phasor data concentrators.

Phasor measurement units and phasor data concentrators tested were found to be largely related to protection relays and other intelligent electronic devices from the same companies. Many of the devices shared the same operating systems, the same software clients for configuration, and the same security approaches. As such the security features available on phasor measurement units and phasor data concentrators was more evolutionary than revolutionary. In many cases this led to less than optimal security solutions such as a reliance on passwords to govern access, a lack of communication integrity for network protocols, and an inability to easily disable unused ports and services at the device. Newer phasor measurement units and phasor

data concentrators show more promise with new features including communication integrity in the form of the new IEC 61850 90-5 protocol and the use of Remote Authentication Dial In User Service (RADIUS) in place of passwords for access control.

References

Common Cyber Security Vulnerabilities Observed in Control System Assessments by the INL NSTB Program. Idaho National Laboratory. Idaho Falls, Idaho 83415, Nov 2008. http://www.inl.gov/scada/publications/d/inl_nstb_common_vulnerabilities.pdf

Control Systems Security Program. National Cyber Security Division. Department of Homeland Security. Catalog of Control Systems Security: Recommendations for Standards Developers, Sept 2009. http://www.us-cert.gov/control_systems/pdf/Catalog_of_Control_Systems_Security_Recommendations.pdf

Department of Homeland Security: Cyber Security Procurement Language for Control Systems (2009). http://ics-cert.us-cert.gov/sites/default/files/FINAL-Procurement_Language_Rev4_100809_0.pdf

File Transfer Protocol (1985) RFC 959. http://tools.ietf.org/html/rfc959

Fink R. Spencer D, Wells R (2006) Lessons learned from cyber security assessments of SCADA and energy management systems. Idaho National Laboratory. Idaho Falls, Idaho 83415, Sept 2006

Fovino IN, Masera M, Guidi L, Carpi G (2010) An experimental platform for assessing SCADA vulnerabilities and countermeasures in power plants. In: 2010 3rd Conference on Human System Interactions (HSI), pp 679–686

Spirent. http://www.spirent.com/

National Institute of Standards and Technology (2009) Recommended Security Controls for Federal Information Systems and Organizations. NIST Special Publication 800–53 Revision 3. http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final.pdf

NMAP Security Scanner. Available at http://www.nmap.org

North American Electric Reliability Corporation (2009) Critical Infrastructure Protection (CIP). NERC StandardsCIP-002-3 through CIP-009-3. http://www.nerc.com/pa/Stand/Pages/CIPStandards.aspx

Open Vulnerability Assessment System (OpenVAS). http://www.openvas.org/

Quick Draw SCADA IDS. http://www.digitalbond.com/tools/quickdraw/

Skaggs B, Blackburn B, Manes G, Shenoi S (2002) Network vulnerability analysis. In: The 2002 45th Midwest symposium on circuits and systems. MWSCAS-2002, vol 3, pp III- 493–495, 4–7 Aug 2002

Stewart J, Maufer T, Smith R, Anderson C, Ersonmez E (2010) Synchrophasor Security Practices. https://www.selinc.com/WorkArea/DownloadAsset.aspx?id=8502

The smart grid interoperability panel—cyber security working group (2010). In: NISTIR 7628 guidelines for smart grid cyber security, vol 2, security architecture and security requirements. Available online: http://csrc.nist.gov/publications/nistir/ir7628/nistir-7628_vol2.pdf

The Syslog Protocol (2009) RFC 5424. http://tools.ietf.org/html/rfc5424

White Paper: Substation Automation for the Smart Grid (2010). http://www.cisco.com/en/US/prod/collateral/routers/ps10967/ps10977/white_paper_c11_603566.pdf

Wireless Procurement Language in Support of Advanced Metering Infrastructure Security. Idaho National Laboratory. Idaho Falls, Idaho 83415, Aug 2009. http://www.inl.gov/scada/publications/d/inl-ext-09-15658_ami_proc_language.pdf

Wurldtech Security Technologies Inc. http://www.wurldtech.com/

Infrastructure Security for Smart Electric Grids: A Survey

Naran M. Pindoriya, Dipankar Dasgupta, Dipti Srinivasan and Marco Carvalho

Abstract The deployment of smart grid technologies is drawing significant attention in the electric power industry. The term "smart grid" refers to modernization of the electric grid using digital technology that includes an advanced sensing and metering infrastructure, high speed, fully integrated two-way communications, and a supporting information infrastructure. In particular, the smart grid combines matured electric grid infrastructure with information and communication technology to offer better grid performance in terms of overall system efficiency and reliability. It supports a two-way energy and information flow, facilitates the integration of time-varying energy sources and new dynamic loads, and amongst other things. However, along with these potential benefits, smart grid also brings new challenges in ensuring security of the grid and the information infrastructure that connects and controls it. This chapter presents a brief survey of various essential components of smart grid and some of the security challenges that encompass virtually all aspects of its operation.

Keywords Smart grids · Smart grids security · Survey · Smart grids architecture · Advanced metering infrastructure

N. M. Pindoriya · D. Srinivasan
Department of Electrical and Computer Engineering,
National University of Singapore (NUS), Singapore 117576, Singapore

D. Dasgupta
Department of Computer Science, University of Memphis, Memphis, TN 38152-3240, USA

M. Carvalho (✉)
Florida Institute of Technology, Melbourne, FL 32901, USA
e-mail: mcarvalho@fit.edu

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_8,

1 Introduction

In electric power industry, the primary focus has always been in maintaining safety and reliability of the grid infrastructure when deploying new power-related equipment. Traditionally the communication and information technologies are considered of secondary importance, often seen as just another device to help achieve power system reliability. However, with the increasing adoption of smart grid technologies, the information infrastructure is becoming critical to the operation and the reliability of the power system. Equipped with this modernized full digital technology and smarter features to handle such complexity, the traditional electric grid is elevated as or referred as smart grid. It transforms the entire electricity value chain, the way electricity is transmitted, distributed, consumed, and charged. A reliable, resilient, secure, flexible, and manageable standards-based information and communication network is the backbone of a smart grid (Pothamsetty and Malik 2009).

Communication network provides the intelligent link between the major elements or domains across entire electrify delivery system. This means distributed intelligent information processing is needed for critical decision-making and performance control, both locally and globally at the entire grid. Adding intelligence throughout the newly networked grid presents many benefits, including improved electric grid reliability and power quality; improved responsiveness; increased transmission and distribution efficiency; and potentially reduced costs for the service providers and customers. It also provides the communication platform for new applications; and builds a suitable economy that ensures future prosperity. Such benefits of reliability and efficiencies in electric grid can be established with the centralized energy and information management, smart devices and applications that enable a finer level of visibility, control and automation.

The key attributes of the smart grid are (U.S. Department of Energy 2009a): the deployment of technologies to enable customers in active involvement; the support distributed power generation and storage; the provisioning of stable power quality; the optimization of power generation, distribution, and consumption. It is also expected that smart grid should anticipate automatic response to system disturbances and load fluctuations; and operate resiliently against physical, cyber attacks, and natural disasters and exhibit self-healing ability.

However, because of increasing adoption of distributed intelligence and broadband communications would also add a new layer of complexity and effectively introduce new challenges. A prerequisite for a safer and secure smart grid is the interoperability of security controls and compliance with standards and regulations. Some of the main security challenges related to smart grid technologies, standards, regulations and management are discussed in this chapter.

2 Overall Architecture of Smart Grid

The smart grid emerges through the integration of advanced information and communication technologies (ICT) into the entire electricity value chain—right from generation to the end users. Dynamic two-way digital communication is possible through ICT at all levels of power grid. The conceptual model of smart grid by the U.S. National Institute of Standards and Technology (NIST), as shown in Fig. 1, defines seven important domains (NIST 2010): bulk generation, transmission, distribution, customers, operations, markets and service providers, along with all the intercommunications and energy/electricity flows among each domain.

Each domain is comprised of important smart grid components connected to each other through two-way communications and energy/electricity paths. The architecture overview based on the three layers—physical power/energy, communication, and application is depicted in Fig. 2. The end-to-end architecture with more details is represented in Leeds (2009).

There are several key elements in detailed logical model such as networks (wired and wireless), functional subsystems (such as the supervisory control and data acquisition (SCADA), etc.), endpoints (e.g., computers in the back offices, monitored and/or controllable substation devices), and overlays (such as distributed security functions and elements) (Metke and Ekl 2010).

The smart grid can be represented as a network of many systems and subsystems, as well as a network of networks, where each domain is expanded into three smart grid foundational layers:

- Physical power/energy layer (generation, transmission, substation, distribution systems, and energy consumer/end users)
- Communication layer (data transport, communications infrastructure and networks)

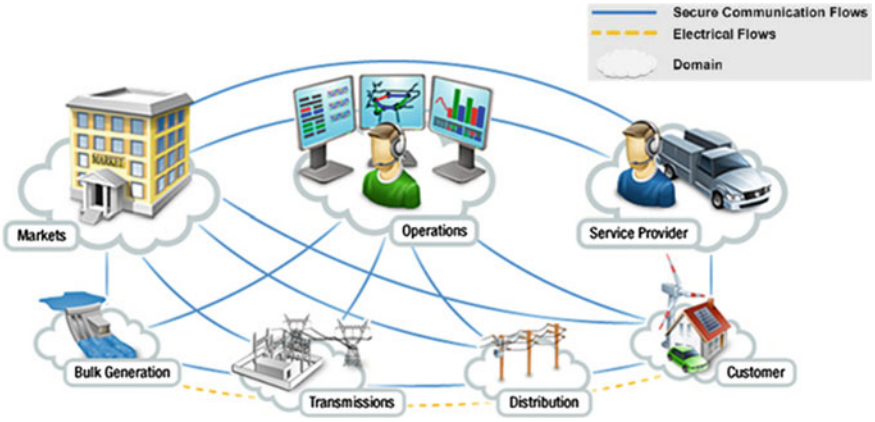

Fig. 1 Smart grid conceptual model by NIST

- Application layer (applications e.g., demand response control, billing, outage control, load monitoring, real-time energy markets, and a new range of customer services).

Of these three layers, the communication layer is the new enabling infrastructure of the power/energy layer that makes the electric grid "smarter". It interconnects a variety of systems and critical devices (such as smart meters, sensors, grid components, and energy management systems (EMS)), physical power/energy layer to application layer, to allow every part of the grid to communicate both up and down the energy supply chain.

The communication layer leverages the same internet technologies that have transformed other high-tech industries. These internet technologies are now enabling distributed intelligent systems to be deployed in the electric grids for many purposes, including remotely monitoring, control, analysis, reporting, forecasting, recovery, and others.

Moreover, to help with the development of required standards, the power industry is gradually adopting different technologies (or rather networks) for the partitioning of the communication layer of the smart grid. Devices and applications in each domain are network end points.

Examples of applications and devices in the customer domain include smart meters, smart appliances, smart thermostats, energy storage, plug-in hybrid electric vehicles (PHEV), and distributed generations (e.g., solar energy as photovoltaic (PV), wind turbines, etc.). Applications and devices in the transmission or distribution domain include phasor measurement units (PMUs) in a transmission line substation, substation controllers, distributed generation, and energy storage.

Applications and devices in the operations' domain include SCADA systems and computers or smart screen at the operations centre. Applications in the operations, market, and service provider domains are similar to those in Web and business information processing. The following technology solutions must be developed and implemented to achieve the vision of the smart grid (NETL 2010):

- Information and communications networks
- Advanced metering infrastructure (AMI)
- Customer side systems and demand response (DR)
- Distribution management system/distribution automation (DMS)
- Transmission enhancement applications and grid optimization
- Distributed energy resources (DER) and storage.

2.1 Information and Communication Networks

On the basis of Smart Electric Grid functional requirements, the network should provide the capability to enable an application in a particular domain to communicate with an application in any other domain over the information network, with proper management control as to who and where applications can be interconnected.

Therefore, the key element of the smart grid is the installation of a completely new two-way information and communication network between the energy suppliers and their customers. This network can be constructed by employing various architectures, with one of the most common being local concentrators that collect data from groups of meters and transmit that data to a central server via a backhaul channel. A variety of communication media and technologies that can be considered to provide part or all of this architecture are includes the copper wiring or optical fiber, hybrid fiber coax, power line carrier (PLC) technology, broadband over power lines (BPL), wireless technologies, internet, etc.

These networks include the Enterprise Network that connects control center applications to markets, generators, and with each other. Local area networks (LAN) are used to identify the network of integrated smart meters, field components, and gateways that form the logical network between distribution substations and a customer's premises; wide area network (WAN) connect the network of upstream utility assets, including-but not limited to-power plants, distributed storage, substations, and so on.

Field area networks (FAN) connect devices, such as intelligent electronic devices (IEDs) that control circuit breakers and transformers; and, home area network (HAN) enable smart appliances, and ultimately smart homes and buildings; as well as the advanced metering infrastructure (AMI) and feedback on demand response. As Fig. 2 shows, the interface between the WAN and LAN consists of substation gateways, while the interface between LAN and HAN is provided by smart meters. These networks may be implemented using public (e.g., the Internet) and non-public networks in combination. Both public and non-public networks will require implementation and maintenance of appropriate security and access control to support the smart grid (NIST 2010).

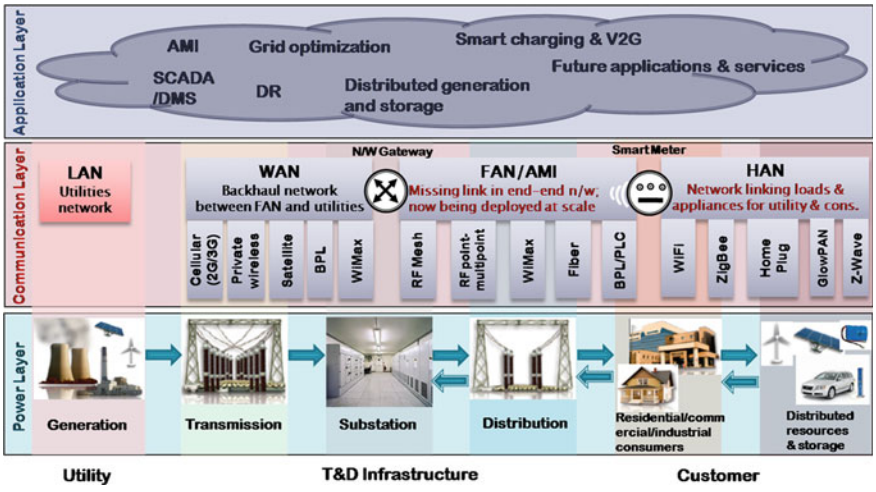

Fig. 2 Simplified three-layer architecture model for smart grid (Leeds 2009)

Low speed transmission SCADA and EMS applications have been successfully integrated among regional transmission organizations, generators, and transmission providers. But these applications need further improvement to fully utilize the integrated, high-speed communications system required by the smart grid. PLC technology has been in use for many years in utility industry. Originally focused on the internet access and voice over internet protocol (VOIP) for consumers, BPL carrier technologies is becoming increasingly accepted and successfully deployed to meet utility needs for DER, automated meter reading (AMR), DR, and consumer portal applications, as well as video monitoring (primarily for security), and other high-speed data applications. Wireless technologies are also currently being developed and demonstrated, but they are not yet used in the grid communication infrastructure on either the system or the user side.

BPL is a technology that allows data to be transmitted over utility power lines. BPL has been implemented in the U.S. and other countries on medium voltage distribution lines, but it has not been applied to HV lines. BPL signals have less attenuation on HV lines and can, therefore, travel longer distance. These are also more amenable to noise-mitigation techniques than distribution lines.

2.2 Advanced Metering Infrastructure (AMI)

One of the important components of AMI is the smart meter. Earlier meters are being replaced by their smarter counterparts, which provide many more capabilities, such as taking usage data in shorter intervals, sending meter readings back to the billing system, alerting in case of power failure, monitoring power quality etc. Public networks (such as the Internet) enabling communications between the service provider and consumers is an integral part of the AMI. Smart meters are envisioned to reside on customer site and send information via the data network to the utility company. Conversely, the utility company can also use the same infrastructure to send commands, updates or new configurations to the remote meters.

An important AMI building block is the data reception and management system (FERC 2013), where the data is collected, processed, and made accessible to the service provider. Integrated home energy management software and related display units simplify user interaction with the AMI system.

An AMI system requires significantly greater bandwidth than automated meter reading (AMR) systems and is more sophisticated than older AMR standards. Besides monitoring, AMI offers efficient control and improved visibility. Energy consumer can utilize smart power sockets and smart appliances to better control consumption, in coordination with the smart meter.

For example, using pool pumps, charging PHEV, heating water, etc. require high power consumption. If many users choose to perform these activities at the peak hour, it may compromise the functionality of the system. AMI components enable users to make smart and mutually beneficial decisions with the utility company, and provide users with the means to actively manage their own consumption. AMI helps the

customer to take advantage of real-time pricing, off-peak rates and other programs offered by utility companies. AMI system can render a host of information including tamper notification, swell events, peak KW readings etc.

2.3 Transmission and Distribution Management and Control

The transmission systems and the centralized generating assets are required to be enable a more reliable and efficient infrastructure. In Smart Grid, we can have wide area monitoring systems (WAMS) that will make use of integrated advanced sensors and phasor measurement units (PMU). WAMS will enable the development and deployment of advanced Grid applications for improved situation awareness. It will also facilitate the system operators' to respond immediately and accurately to any disturbance in the system (U.S. Department of Energy 2009b). Development of advanced operational algorithms and compatible controlling mechanisms may enable the design of resilient and seal-healing capabilities for the Grid.

The North American Synchro Phasor Initiative (NASPI) organization has undertaken a PMU installation roadmap, and standardization of voltage levels for PMUs. PMU is used to read voltages, currents, frequency and other information with very high accuracy, speed and synchronization. These measurements are also known as synchro phasors. Measurements are highly synchronized and time-aligned by means of accurate GPS- synchronized clocks. A more accurate picture of the large grid system can be formed, for example, by combining measurements from various PMU sites.

Applications can process these high dimensional data read from various points and provide grid operators with features to visualize the state of the grid in a comprehensive manner and with high fidelity. These applications will also be able to aid experts to notice evidence of changing conditions and nascent grid anomalies. With sophisticated data visualization and analysis software, these new measurements can be used to drive power system planning and forensic analysis (U.S. Department of Energy 2009b).

2.4 Distributed Energy Resources and Energy Storage

Smart grid will allow integration of various sources such as wind turbines, micro turbines, photovoltaic (PV) arrays, etc. Plug-in hybrid electric vehicle (PHEV) could also act as a temporary energy source if necessary. All these sources are time-varying and distributed in nature. These will help the utility to a great extent to satisfy dynamic load, especially at the peak hour. Enhanced communication and automation is required for large-scale deployment of DER and to integrate these time-varying sources into their grid without creating instability and load balancing problems.

To render the energy demand instantaneously and in a stable manner, the DER integration will require backup source for generation and energy storage system. All these will help the utility to meet the peak demand without investing huge amount of resource for a comparatively very short time period. This will result in reduced energy cost for customers and encourage utility companies to help and reward cooperating parties involved in generation and charging/discharging activities. It will also enhance the integration of non-dispatchable generation resources into the power grid. Besides PHEV, other existing options for energy storage are batteries, pumped storage hydroelectric, compressed air storage, flywheel storage, thermal storage, and magnetic superconducting storage.

3 Smart-Grid Standards and Interoperability

In order to realize the promise of smart grid, it is critically important to define and adopt an appropriate set of standards, as they directly affect the interoperability, compatibility, reliability, and efficiency of the overall system. Standards define specifications for languages, communication protocols, data formats, linkages within and across systems, interfaces between software applications and between hardware devices, and much more. One key requirement of the smart grid is the interoperability of the cyber systems used to manage the power systems. Interoperability can be defined as the ability of two or more networks, systems, devices, applications, or components to exchange and readily use information—securely, effectively, and with little or no inconvenience to the user (IEEE 1990).

The Smart Grid will be a system of interoperable systems. Standards are essential for each interface to support many different Smart Electric Grid applications and also needed for data networking and cyber security throughout the grid. Security standards are used to establish requirements on the security operations of energy service providers (e.g., utilities, generators, system operator, etc.) as well as smart grid device manufacturers. Many power industry standards organizations, for example, the National Institute of Standards and Technology (NIST), the North American Electrical Reliability Corporation (NERC), the International Society of Automation (ISA), the National Infrastructure Protection Plan (NIPP), and the American National Standards Institute (ANSI), are assisting in the development of frameworks for both smart grid standards and security requirements. The NIST has primary responsibility for coordinating development of an interoperability framework allowing smart grid technologies to communicate and work together. The Electric Power Research Institute (EPRI), Institute of Electrical and Electronics Engineers (IEEE), and International Electrotechnical Commission (IEC) are also working alongside NIST to develop guidelines for smart grid security and interoperability. The NIST has identified five "foundational" sets of standards for smart grid interoperability and cyber security that are ready for consideration by federal and state energy regulators. Over 100 standards have been identified by IEC, focus on information models and protocols critical to reliable and efficient grid operations as well as cyber security.

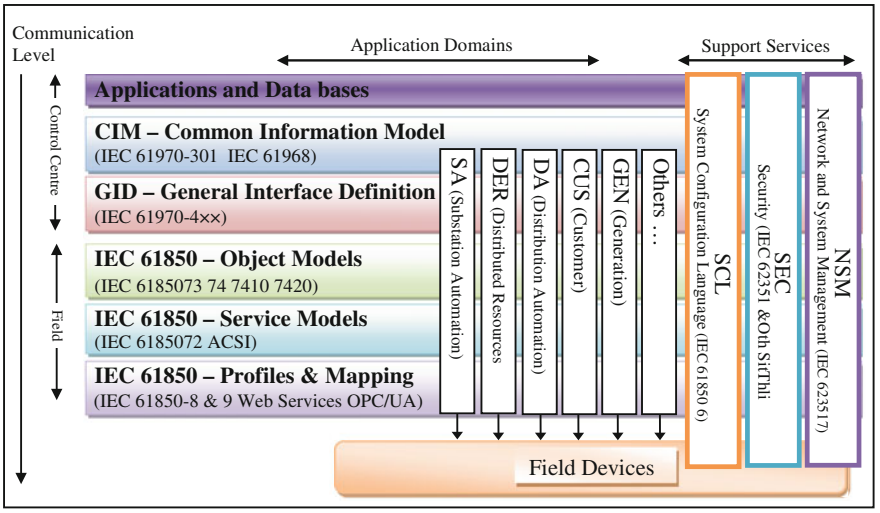

Fig. 3 IEC 61850 Models and Common Information Model (CIM) (CSWG 2010)

These standards will be further updated to achieve efficient and secure intersystem communications as Smart Electric Grid requirements and technologies evolve. The core IEC standards and their functions are given below (IEC 2013). The specialized communication standards developed by IEC are illustrated in Fig. 3.

- IEC 61970 and IEC 61968: Providing a common information model (CIM) necessary for exchanges of data between devices and networks, primarily in the transmission (IEC 61970) and distribution (IEC 61968) domains. CIM standards are integral to the deployment of a smart grid scenario, in which many devices connect to a single network.
- IEC 61850: Facilitating substation automation, distributed generation (solar PV, wind power, fuel cells, etc.), SCADA communication and distribution automation as well as interoperability through a common data format.
- IEC 60870-6: Facilitating exchanges of information between control centers.
- IEC 62351: Addressing the cyber security of the communication protocols defined by the preceding IEC standards. Cyber security is such a major concern with smart grids, which are especially vulnerable to attack because of the two-way communication between devices and the utility grid.

These five IEC standards were among the 25 smart grid-relevant standards identified as "ready for implementation" in the *NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0*, which was issued in January 2010 (NIST 2010).

However, these specifications required a cyber security review that could not be completed until NIST finalized its initial *Guidelines for Smart Grid Cyber Security*, which were published in early September 2010 (CSWG 2010). CIM enables vertical

and lateral integration of applications and functions within the Smart Grid. IEC 61850 and its associate standards are emerging as favorites for WAN data communication, supporting TCP/IP, among other protocols, over fiber or a 1.8-GHz favor of WiMax.

The communication and interoperability standards developed by IEEE (2013) are IEEE 802.3 (Ethernet), IEEE 802.11 (WiFi), IEEE 802.15.1 (Bluetooth), IEEE 802.15.4 (Zigbee), IEEE 802.16 (WiMax). Moreover, the most relevant ANSI standards for interoperability of AMI systems include ANSI C12.19 (metering "tables" internal to the meter) and ANSI C12.22 (communications for metering tables). ANSI C12.22 and its associate standards are viewed as the favorite LAN standard, enabling a new generation of smart meters capable of communicating with their peers as well as with their corresponding substation gateways over a variety of wireless technology.

NIST has developed very important guidelines related to cyber security information technology and secure interoperability. Two documents of particular interest for the Smart Electric Grid are: NIST SP-800-53 (recommended security controls for federal information systems) and NIST SP-800-82 (Guide to industrial control systems (ICS) Security). Broadly speaking, there are four levels of cyber security standards (Cleveland et al. 2008):

1. Media related standards—Specific to fiber optics, microwave, WiFi, wires, telephones and cellphones
2. Transport related standards—Internet standards including Ethernet, Internet Protocol (IP), Transport Control Protocol (TCP), Hyper Text Transfer Protocol (HTTP)
3. Application related standards—Hyper Text Markup Language (HTML), EXtensible Markup Language (XML), IEC 61850, Common Information Model (CIM)
4. Security related standards—Advanced Encryption Standard (AES 256), Public Key Infrastructure (PKI), secret keys, and certificates.

4 Smart Grid Security and Cyber Security

The Smart Electric Grid will require the development and deployment of several new technologies, such as smart meters, sensors, extensive computer network connectivity for a two-way communication infrastructure that supports the new sophisticated features like real-time information and control for electric power grid. However, such increasing network connectivity and telecommunication capabilities, in smart grid, require both physical security and cyber security management to safeguard the critical infrastructure elements across entire electricity delivery system. Cyber security, in particular, has always been a concern for utility IT experts, but has become a more significant issue due to the increasing penetration of smart grid technology. These technologies expand the application of network information systems to utility and customer assets that previously required manual operation and were not remotely accessible. Securing the assets of electric power delivery systems, from the control centers to the substations, to the feeders and even to customer meters, require an

end-to-end security infrastructure that protects the large number of communication assets (control center-based SCADA, RTUs, PLCs, power meters, digital relays, and bay controls) used to operate, monitor, and control power flow and measurement. Security of energy systems and electronic information in the IT and telecommunication infrastructure includes the confidentiality, integrity, and availability on all related cyber physical systems.

- **Integrity** of telemetry data and control commands is the most critical security requirement for the proper functioning of the smart grid and support the consisting and accuracy in delivery and billing. It includes assurance that data has not been modified without authorization, source of data is authenticated, timestamp associated with the data is known and authenticated, and quality of data is known and authenticated.
- **Availability** is generally considered the next most critical security requirement, although the time latency associated with availability can vary 4 ms for protective relaying, sub-seconds for transmission wide-area situational awareness monitoring, seconds for substation and feeder SCADA data, minutes for monitoring non-critical equipment and some market pricing information, hours for meter reading and longer term market pricing information, and days/weeks/months for collecting long term data such as power quality information.
- As the smart grid reaches into homes and businesses, and as customers increasingly participate in managing their energy and their information is more easily available in cyber website, **confidentiality** and **privacy** of their information have increasingly become a concern. Unlike power system reliability, customer privacy is a new issue. Not only this, electric market information and general corporate information, such as human resources, internal decision-making, etc. have some confidential proprietary value.

A lack of adequate security in the electric power industry could pose threats of service disruption, which can impede safe and efficient functioning of the electric grid. The most common malicious attack includes both attempts to physically tamper with a meter, and disruption of the supporting communications infrastructure. Security is not just about preventing attacks or system compromises; it is also about preventing the accidental or malicious leakage of information. Particularly for real-time operations, it is crucial to "live through" any attacks or compromises to the information infrastructure, and to recover with minimal disruption to the power system operations—including power system reliability, efficiency, and cost.

Because of the large variety of communication methods and performance characteristics, as well as because no single security measure can counter all types of threats, it is expected that multiple levels of security measures needs to be implemented. Added layers of security controls, policies, and procedures are necessary to prevent, detect and manage security risks in power grids.

Security of information exchanges will also require end-to-end security from the sender of the information all the way across through all intermediate paths to the final receiver of the information. Accordingly, information security must address not only fixed assets and devices but also the virtual paths and information flow from

end to end. This end-to-end aspect of security makes it far more difficult to assess the control risks—threats, vulnerabilities, and impacts—as well as to determine the most appropriate security solutions. The NIST smart grid cyber security coordination task group (CSCTG) was established to ensure consistency in the cyber security requirements across all the smart grid domains and components.

Thus the most critical infrastructure necessary to create a reliable high-performance smart grid is the information and communications networks. Functional requirement of communication infrastructure is that the network should enable an application in a particular domain to communicate with an application in any other domain in the information network, with proper management and control over who and where applications can be interconnected.

4.1 AMI Security

The operational imperatives for AMI Cyber Security (AMI-SEC) Task Force (2008) recognize the existence of gaps in risk management between AMI and traditional information and communications technology (ICT) systems. Typically, AMI lies at the intersection of physical and logical infrastructures. AMI's resiliency not only demands security and continuity, but rethinking the relationship of systems to services. Without proper security in AMI systems, electricity distribution will be unreliable and interruptible both on a physical and logical scale. An AMI system's potential exposures may exist in control functions in the form of remote service disconnects and management of devices in home area networks (HAN). These potential exposures exemplify the increased risk against the grid as a whole. AMI-SEC Task Force (2008) was developed the security domain model to boundary the complexity of specifying the security requirement to implement a robust, secure AMI solution as well as serve as a tool to guide utility companies in their AMI implementation. The "services" provided in Table 1 are described for security domains (AMI-SEC-ASAP 2008).

4.2 Substation Security

Substations, transmission and distribution domains include the devices such as circuit breakers, power transformers, phase-shifting transformers, capacitor banks, switches, etc. It may also contain various electronic automation and communication devices used to measure, monitor, and control the substation components.

The increasing level of automation envision in smart grids, are likely may open the door for malicious activities. In other words increased automation, if not performed with comprehensive security evaluation, can result in new vulnerabilities related to the substation devices. Potential consequences of exploitation of vulnerabilities

Table 1 Various applications and associated services

Security domain	Description
Utility Edge Services	All field services including monitoring, measurement and control to be controlled by the Utility provider
Premise Edge Services	All field service applications including monitoring, measurement and control controlled by the customer (customer has control to delegate to third party)
Communications Services	These applications relay, route, and perform field aggregation, field communication aggregation, field communication management information
Management Services	Provide support services for automated and communication services (includes device management)
Automated Services	Allow unattended collection, transmission of data and performs the necessary translation, transformation, response, and data staging
Business Services	Support core business applications (includes asset management)

resulting from substation automation are grid instability, power outages, destruction of generators, which are highlighted in U.S. Department of Energy (2009b).

To automate substation without introducing new security vulnerabilities, and important security goals should be given high priority. Authentication and authorization must be enforced to prevent intruders and unauthorized operator users from accessing and tampering with distribution devices, running unauthorized commands.

The transformation must also ensure integrity and confidentiality of telemetry data, control protocols and other administrative information. It must also emphasize the protection of upstream assets. The processes used to manage energy routing from plant to consumer and fidelity of the energy delivery systems are defined by the telemetry and control systems security zone.

4.3 SCADA Security

SCADA systems now are becoming increasingly connected to the public Internet, which enables the fielding of new equipments such as intelligent electronic devices (IEDs). Such deployments, however, are likely to make the SCADA system more vulnerable (Ericsson 2010). Remote device monitoring is the key to enhancing the reliability of the electric power grid. The substations along the electricity supply chain contain many remote terminal units (RTUs) and IEDs.

However, the monitored data from these substations can only be relied upon if the integrity of the data is assured. The various new issues related to SCADA system have been emphasized in the CIGRÉ working groups JWG D2/B3/C2.01 “Security for Information Systems and Intranets in Electric Power Systems” (Ericsson et al. 2007) and D2.22 “Treatment of Information Security for Electric Power Systems” (Ericsson et al. 2010).

There is a need to perform case-by-case technical assessments of all vendor products, studying provide expose and the protocols they use so that their security can be assured within the greater cyber security context. Security for such devices is being standardized according to the IEC 62351 standards.

4.4 Communication Link Security

As wireless devices are inexpensive and provide significant advantages over wired counter parts, they are being prevalent in current Smart Grid deployments. Many AMI implementations are using mesh networks because its reliability and redundancy. However, their security aspects are not publicly available at this point and most of these implementations are based on IEEE 802.15.4 protocol and there are known vulnerabilities with its implementations.

4.5 Security Protocol Design Challenges

The incorporation of advanced ICT in power grid will require many communication protocols. It is likely that many of these protocols will be based on customized versions of existing protocols, borrowed from other domains. However, the adaptation and use of technologies from different domains require an appropriate mapping and match at several abstract levels, including the conceptual level, system policy level, formats and algorithms level and implemented tools level. A detailed discussion and case study of such adaptations can be found at Khurana et al. (2007).

If for a particular level or application, requirements for the power grid are significantly unique, new protocol should be designed from scratch. Protocols should also be modular and must provide multiple levels of defense, so that they can be easily replaced with newer modules in case of discovery of a flaw. Computational and communication overhead of the protocol must be evaluated. Error handling is another important area that required significant attention while designing protocols. Protocol must guarantee the authenticity of the messages and try to optimize other issues.

4.6 Cryptographic Key Management

Managing and protecting keys is a problematic task for many users. Cryptographic functions are computationally expensive, too, especially for smart devices with limited physical resources. So, we can follow the approach of Computational Grid (C-Grid), to some extent. Still, for power grid (P-Grid), much customization is required. Federated authentication services like single sign-on may also be used.

4.7 Error Handling Challenges

Systems that do not perform error checking of invalid inputs are susceptible to crush and execute arbitrary code at some point. On the other hand rigorous bounds checking and error management can sometimes lead to problems, too. So, balance must be maintained so that the system does not execute alien code and can survive from distributed DoS attacks.

4.8 Challenges due to Ad Hoc Automation

Although the envisioned Smart Grid deployment has not been completed yet, the current grid is being automated and updated in different parts on a regular basis. New smart meters are being deployed for operational conveniences. Many of these improvements are being carried out without any comprehensive security assessment. These smart devices have lifecycle of several years. Competition to be the first to market, many of the vendors are rushing and not providing considerable care in comprehensive security consideration.

4.9 Smart Grid Security Challenges: Summary

There are many security challenges need to be addressed as the power grid is fully integrated to the cyber infrastructure, we summarize some important smart grid security challenges in Table 2.

5 Conclusions

The basic framework of Smart Electric Grid and most significant technologies/ elements that must be developed and implemented to achieve the vision of the Smart Electric Grid are briefly discussed in this chapter. To achieve efficient and secure intersystem communications and to properly manage and control the power system, one key requirement of the Smart Electric Grid is the standards related to interoperability and cyber security.

Various organizations are actively involved in the development of these important standards. This chapter briefly surveyed the different standards for Smart Electric Grid. Moreover, one critical aspect of the Smart Electric Grid related information and communications infrastructure is the physical and cyber security. Infrastructure security that includes the protection of networks and servers from unauthorized accesses and malicious attacks are discussed.

Table 2 Summary of important smart grid security challenges

Specific problem	Description of problem
Privacy of usage data (McDaniel and McLaughlin 2009; IEEE 2010a)	• Usage data and statistics will be of paramount importance to entities for business intelligence and also to malicious hackers for launching targeted and well informed attacks. Existing policies do not describe all possible scenarios • Government need to establish a national regimen for consumer protections
Vulnerabilities and exploitation restriction (McDaniel and McLaughlin 2009; IEEE 2010b; INL 2009)	• General-purpose operating systems doing real time control is far riskier than well tested finite state machines implemented in special purpose hardware. Limits must be imposed on systems to limit worst-case behaviors. These limit conditions must be locally controlled and must not be remotely programmable • Hackers may gain unauthorized access to Smart Meters to manipulate its functionalities and make the grid unstable and cause financial loss • Extraction of data and encryption key from device memory was demonstrated
Communication networks and links vulnerability (IEEE 2010a; INL 2009)	• Most mesh networking protocols are based on IEEE 802.15.4 standard which is reported to be susceptible to a set of known types of attacks • As the vendors of wireless AMI technology are in a rush to develop and deploy market, security may not receive sufficient emphasis
Key management (IEEE 2010a; Khurana et al. 2007; Sugwon and Myongho 2010)	• Currently, it's not feasible to operate Cryptographic key management or similar services for 5.5M smart meters. Key management is required to be planned for various communication modes, such as master-to-IED, peer-to-peer and broadcast • Analogous C-Grid tools can be used in P-Grid to facilitate single sign-on based access to a multitude of resources across organizational boundaries. Scaling such accesses requires federated approaches where the organizations agree on a common authentication and authorization system for accessing data and resources
Intrusion detection (IEEE 2010b; Khurana et al. 2007; Sugwon and Myongho 2010)	• Organizations need to define security policies, deploy monitoring systems supported by advanced IDSs, and set up mechanisms for forensic analyses and development of a communication process to share incident data with other organizations and a response in a coordinated fashion
Operational cost (IEEE 2010b)	• Operational cost of a million-node network, where significant portion of resources need to be invested in monitoring and continuous analyzing of threats and compromises, is another challenging issue

(continued)

Table 2 (continued)

Specific problem	Description of problem
Development cost effectiveness (Sugwon and Myongho 2010)	• Implementing complex security functionalities on embedded microprocessor based platform is not feasible because of these devices' limited computational ability. Besides security functions, these devices also have to receive network packets and are vulnerable to Kernel live-lock resulting from too much interrupt handling. For capturing high-rate arriving packets, an expensive alternative is to use specialized hardware such as network processors in the monitoring cards. Some dedicated hardware, possibly FPGA-based co-processors or hardware accelerators or graphic processor may be used to for cryptographic functionalities. One less expensive alternative is to use Chip-level MultiThreading processor, such as ARM11 MPCode (quad-core)
Transition (Sugwon and Myongho 2010; Kouril et al. 2006)	• Transition is another challenging task involving finding workarounds for already deployed IEDs with limited processing capabilities and lack of security features. For such devices, there are two transition scenarios for achieving security: "bump-in-the wire (BITW)" and "bump-in-the stack (BITS)" solutions • Transition cost will also be very high for setting up smart metering infrastructure. But some interesting findings suggest that end consumers are willing to make onetime payment of $48 on average or $13 per month. This will offset the transition expense to great extent
Error handling (IEEE 2010b; Khurana et al. 2010)	• Only the input sequences that are within defined safety boundaries should be allowed • Common techniques include back off timers, limits on number of events reported, event reporting compression and suppression techniques, and both in-band and out-of-band reporting. Increased complexity in the error management process, leads to an increase in edge cases as well • Protocols having no mechanism to handle malformed or unexpected packets may fail or possibly execute arbitrary code
Fault and failure modeling (IEEE 2010b)	• Our current modeling capability is also limited. We need to model faults and failures more accurately with respect to security incidents. We must also undertake analyses assuming a substantial number of faults and failures in wide array of combinations
Recovery after failure (McDaniel and McLaughlin 2009; IEEE 2010b)	• Comprehensive recovery strategies must also be developed through close collaboration between utility companies and vendors • A backup plan is necessary that will allow some level of power operations when the computers don't work properly

(continued)

Table 2 (continued)

Specific problem	Description of problem
Efficiency (IEEE 2010a)	• Efficiency and scalability must meet varying real-time requirements based on the location. Constrained network and devices must also be considered
Evolvability (IEEE 2010a; Khurana et al. 2010)	• The design should be modular, so that it can be easily upgraded later on with minimal disturbance to other components working properly
Data management (Khurana et al. 2007)	• Managing and accessing large amount of energy usage and business related data at control center are challenging issues. To utilize this data several C-Grid techniques such as data federation, data virtualization and integration and data location services can be realized
Load management (Khurana et al. 2007)	• Energy can be treated as a resource and its delivery to an appliance can be treated like a task. Now we can utilize C-Grid's task scheduling techniques to efficiently perform load management in the P-Grid
Breaching of trust (IEEE 2010a)	• Most of the control systems' design decisions are based on implicit trust. So, methods to deal with untrustworthy participants are required
Security protocol design challenges (Khurana et al. 2010; Sugwon and Myongho 2010)	• Security goal should aim for complete guarantee for message authenticity and integrity from protocols. The grid applications require high performance, high availability, timeliness, comprehensive security design, adaptable and evolvable designs etc. So, protocols must be developed considering these fundamental constraints • For power grid a potential approach is to use existing protocols upon customization for P-Grid. New protocols could be designed, too, if requirements are fairly unique. In Khurana et al. (2010), design principles are proposed considering traditional tools, known cyber attacks and protocol goodness properties • Computation and communication overhead should also be analyzed thoroughly for designing efficient protocols. Error handling, detection of cyber attack & proper responses and evolvability should also be considered carefully to design a highly available protocol. The design should be modular, so that it can be upgraded easily
System complexity (INL 2009)	• Risk resulting from cyber attack depends on threats, vulnerabilities and consequences which is very difficult to estimate. The size and dynamic nature deteriorates the scenario by adding up more complexity and uncertainty. It is hard to predict how an attack can be manifested given the adversary is reasonably intelligent. Emphasis on cost minimization and market capturing than security analyses in the face of incomprehensible threats may result in deployment of vulnerable systems

(continued)

Table 2 (continued)

Specific problem	Description of problem
SCADA automation without proper security (INL 2009)	• SCADA Security assessment has been conducted in the past and several vulnerabilities have been found. Among 17,325 transmission substations in the US and Canada, 81 % and 57 % of distribution substation have some form of automation and these are increasing day by day. Study found that new vulnerabilities are introduced which are associated with substation automation
PMU Vulnerabilities (INL 2009)	• NASPI (North American SynchroPhasor Initiative) security has not yet been sufficiently explored. In Smart Grid there are many physically unprotected potential entry points. Wireless networks can be easily sniffed by attackers and these are susceptible to Man-in-the-Middle attacks. There are weaknesses in security mechanisms to prevent these attacks
Massive deployment of intelligent sensors (Jain and Chapman 2010)	• Large scale deployment of new tiny devices, called Heterogeneous System on a Chip (HSoC), is proposed to sense and autonomously reconfigure power system. Three layered sensor is envisioned. VLSI implementation details of the first two layers, powers system sensing and decision making layers. In addition to detection of failures in power system, handling of chip faults are also addressed
Cooperation among various organizations (Khurana et al. 2007)	• Collaboration and sharing among various organizations is needed to fight against intrusion and malicious activities
Lack of comprehensive security strategy (U.S. Department of Energy 2009a; IEEE 2010b)	• We must develop security architecture appropriate to the Smarter Grid's needs. The power engineering community must welcome security community and work together

References

AMI-SEC-ASAP (2008) AMI-SEC-ASAP, AMI system security requirements-v1.01, December 2008. http://osgug.ucaiug.org/utilisec/amisec/default.aspx. Accessed 15 March 2011

Cleveland F, Small F, Brunetto T (2008) Smart grid: interoperability and standards—an introductory review. Utility Standard Board, September 2008

CSWG (2010) Cyber Security Working Group (CSWG) of the smart grid interoperability panel (SGIP)—NIST. Introduction to NISTIR 7628. Guidelines for smart grid cyber security. September 2010. http://www.nist.gov/smartgrid/upload/nistir-7628_total.pdf

Ericsson GN (2010) Cyber security and power system communication—essential parts of a smart grid infrastructure. IEEE Trans Power Delivery 25(3):1501–1507

Ericsson G, Torkilseng A, Dondossola G, Jansen T, Smith J, Holstein D, Vidrascu A, Weiss J (2007) Security for information systems and intranets in electric power systems. Tech. Brochure (TB) 317 CIGRÉ

Ericsson G, Torkilseng A, Dondossola G, Pietre-Cambacédès L, Duckworth S, Bartels A, Tritschler M, Kropp T, Weiss J, Pellizzonni R (2010) Treatment of information security for electric power utilities (EPUs). Tech. Brochure (TB), CIGRÉ

Federal Energy Regulatory Commission (FERC) (2013), http://www.ferc.gov/eventcalendar/Files/20070423091846-EPRI%20-%20Advanced%20Metering.pdf
IEEE (1990) IEEE standard computer dictionary: a compilation of IEEE standard computer glossaries, New York
IEEE (2010a) Smart grid security issues, Computer and reliability societies, IEEE, January/February 2010
IEEE (2010b) The smarter grid, Computer and reliability societies, IEEE, January/February 2010
IEEE (2013) Approved IEEE smart grid standards. http://smartgrid.ieee.org/standards
INL (2009) INL/EXT-09-15500, Study of security attributes of smart grid systems—current cyber security issues, April 2009. http://www.inl.gov/scada/publications/d/securing_the_smart_grid_current_issues.pdf. Accessed 15 March 2011
International Electrotechnical Commission (IEC) (2013) http://www.iec.ch/
Khurana H, Khan MMH, Welch V (2007) Leveraging computational grid technologies for building a secure and manageable power grid. 40th annual Hawaii international conference on system sciences
Khurana H, Bobba R, Yardley T, Agarwal P, Heine E (2010) Design protocols for power grid cyber-infrastructure authentication protocols
Jain VK, Chapman GH (2010) Massively deployable intelligent sensors for the smart power grid. IEEE 25th international symposium on defect and fault tolerance in VLSI systems (DFT)
Kouril D, Matyska L, Procházka M (2006) Improving security in grids using the smart card technology. Grid computing, 7th IEEE/ACM international conference
Leeds DJ (2009) The smart grid in 2010: market segments, applications and industry players. GTM Research, July 2009
McDaniel P, McLaughlin S (2009) Security and privacy challenges in the smart grid. http://www.patrickmcdaniel.org/pubs/sp-smartgrid09.pdf. Accessed on 15 March 2011
Metke AR, Ekl RL (2010) Security technology for smart grid networks. IEEE Trans Smart Grid 1(1):99–107
National Energy Technology Laboratory (NETL) (2010) Understanding the benefits of the smart grid. Report (DOE/NETL-2010/1413), June 2010
NIST (2010) Report on NIST framework and roadmap for smart grid interoperability standards, release 1.0, January 2010. http://www.nist.gov
Office of the National Coordinator for Smart Grid Interoperability, National Institute of Standards and Technology (2010) NIST framework and roadmap for smart grid interoperability standards. Release 1.0 (NIST SP 1108), January 2010. http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pdf. Accessed 15 March 2011
Pothamsetty V, Malik S (2009) Smart grid leveraging intelligent communications to transform the power infrastructure. CISCO Systems, white paper, February 2009
Sugwon H, Myongho L (2010) Challenges and direction toward secure communication in the SCADA system. Communication networks and services research conference (CNSR)
The North American SynchroPhasor Initiative (NASPI) (2013) www.naspi.org
U.S. Department of Energy (2009a) Smart grid system report, July 2009
U.S. Department of Energy (2009b) Study of security attributes of smart grid systems—current cyber security issues. Report by U.S. Department of Energy (DOE), Office of Electricity Delivery and Energy Reliability, April 2009

Known Secure Sensor Measurements Concept and Its Application for Critical Infrastructure Systems

Annarita Giani, Ondrej Linda, Milos Manic and Miles McQueen

Abstract The manipulation of critical physical processes and the falsification of system state is a relevant concern for many modern control systems. Common approaches to this problem such as network traffic and host based state information analysis feature difficulties such as high false alarm rate. Furthermore, issues in integrating the system state falsification detection into an existing control system such as cost or technical issues, impose additional difficulties. To alleviate these issues, a low cost and low false alarm rate method for improved cyber-state awareness of critical control systems, the Known Secure Sensor Measurements (KSSM) method, was proposed by the authors of this chapter. This chapter reviews the previously developed theoretical KSSM concept and then describes a simulation of the KSSM system. The presented KSSM method constitutes a reliable mechanism for detecting manipulation of critical physical processes and falsification of system state. Unlike other network based approaches, the method utilizes the physical measurements of the process being controlled to detect falsification of state. In addition, the KSSM method can be incrementally integrated with existing control systems for critical infrastructures. To demonstrate the performance and effectiveness in detecting various intrusion scenarios, a simulated experimental control system network was combined with the KSSM components. The KSSM method is intended to be incorporated into the design of new, resilient, cost effective critical infrastructure control systems.

A. Giani (✉)
CNLS, Los Alamos National Laboratory, Los Alamos, USA
e-mail: annarita.giani@gmail.com

O. Linda · M. Manic
University of Idaho, Moscow, USA
e-mail: olinda@uidaho.edu

M. Manic
e-mail: miskomis@gmail.com

M. McQueen
Idaho National Laboratory, Idaho Falls, USA
e-mail: Miles.McQueen@inl.gov

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8_9,

1 Introduction

Resiliency and enhanced state-awareness are crucial properties of modern control systems. Critical infrastructures, such as energy and industrial systems, would benefit from being equipped with intelligent components for timely reporting and understanding of the status of the control system. This goal can be achieved via complex system monitoring, real-time system behavior analysis and timely reporting of the system state to the responsible human operators (Linda et al. 2011a).

In Rieger et al. (2009) a resilient control system was defined as follows: *one that maintains state awareness and an accepted level of operational normalcy in response to disturbances, including threats of an unexpected and malicious nature*. Here, the enhanced state-awareness is understood as a set of diverse performance criteria such as cyber or intelligent analysis that is used to maximize the adaptive capacity of the system to respond to threats.

Falsification of physical system state can pose significant danger to the operation of a control system. An intelligent adversary attempts to deceive the operator with the intention to achieve desired manipulation of the control system without early detection. An intuitive way for achieving this task is modification of physical measurement values sent to the operators by injecting false information. Hence, protection of measurement values is of high importance. There exist cryptographic techniques that provide sufficient level of information protection (Schneier and Ferguson 2010; Stamp 2011). However these techniques require increased computational cycles, increased power, and higher available network bandwidth, which might not be available on many currently deployed control systems.

To address these issues, a novel, low cost, low false alarm rate, and high reliability detection technique for identifying manipulation of critical physical process and falsification of system state was previously proposed (Giani et al. 2011; McQueen and Giani 2011). This technique, called Known Secure Sensor Measurements (KSSM), uses the idea of obtaining a randomly selected subset of encrypted (i.e. known secure) physical measurements that are sent in sequence after the plain-text (i.e. insecure and unencrypted) measurements used for control. The subsequent comparison of the randomly selected plain-text and the known secure values reveals potential system falsification. By randomly modifying this selected subset of KSSM sensors, a complex cyber-state awareness of the control system and falsification of system state can be maintained while imposing as little additional computational and bandwidth cost as desired. Hence, by utilizing the physical measurements themselves for aiding cyber-security, the KSSM method differs from traditional approaches to network system security such as anomaly or signature detection systems (Linda et al. 2009, 2011b; Yang et al. 2006; Zhong et al. 2007).

A variety of techniques for protecting critical infrastructure control systems from cyber attacks have been proposed. These proposals have included cryptographic techniques such as those recommended in AGA-12 (American Gas Association 2005a,b), intrusion detection for industrial control systems (Cheung et al. 2007; Tsang 2005), use of deception (McQueen 2009), and many other general techniques and concepts

for securing IT systems from cyber attack which have been adapted to control systems. While all of these adaptations seem to have some merit for protecting control systems, relatively few have focused on the fundamentally unique aspects of general control systems, which in our view is that they control a physical process; have much longer life cycles than standard IT systems; and may have severe resource constraints, including cost. There are exceptions to this of course, such as the large body of research into protecting the electric power grid, see for example (Giani et al. 2011; Wei 2010). The Known Secure Sensor Measurement technique described in this chapter is unique in that it blends IT security concepts with the physical processes measurement and control in order to enhance the detection of an attack on the physical process even if the entire communication infrastructure has been compromised through a cyber attack.

In this chapter we first explain the KSSM concept, including the technical objectives and research approach and then we will show on specific simulation examples how the KSSM system could be implemented. The overall architecture of the system is presented, followed by description of the two major components, Sensor Selector and Signal Analyzer. The Sensor Selector uses an algorithm to perform pseudo-random sensor selection based on multiple criteria. The Signal Analyzer contains a buffer of requested KSSM values and performs measurement comparison and system state falsification detection. The designed KSSM system architecture was integrated with a virtual control system communication network. The performance of the system is demonstrated on several test scenarios. As a practical application we show how the generalized concept of known secure sensor measurement can be used as a countermeasure against a collection of data integrity attack to the smart grid.

This chapter provides an overview of the previously published work on the KSSM concept (Linda et al. 2011a; McQueen and Giani 2011).

2 Known Secure Sensor Measurement Concept

The concept of Known Secure Sensor Measurements was previously proposed in McQueen and Giani (2011). The KSSM technique constitutes a novel low cost, low false alarm rate, and high reliability detection technique for identifying malicious manipulation of critical physical processes and the associated falsification of system state. The fundamental idea of the method is to obtain a randomly selected subset of encrypted (known secure) physical measurements that are sent in sequence together with the plain-text (unencrypted) measurements used for control. The comparison of the randomly selected plain-text and KSSM values reveals potential falsification of system state.

The developed KSSM concept is targeted for critical infrastructure control systems that lack robust cryptographic techniques and have limited computational and communication bandwidth resources. It is important to note here that most critical infrastructures fit well within this targeted group. Hence, the KSSM method is widely applicable.

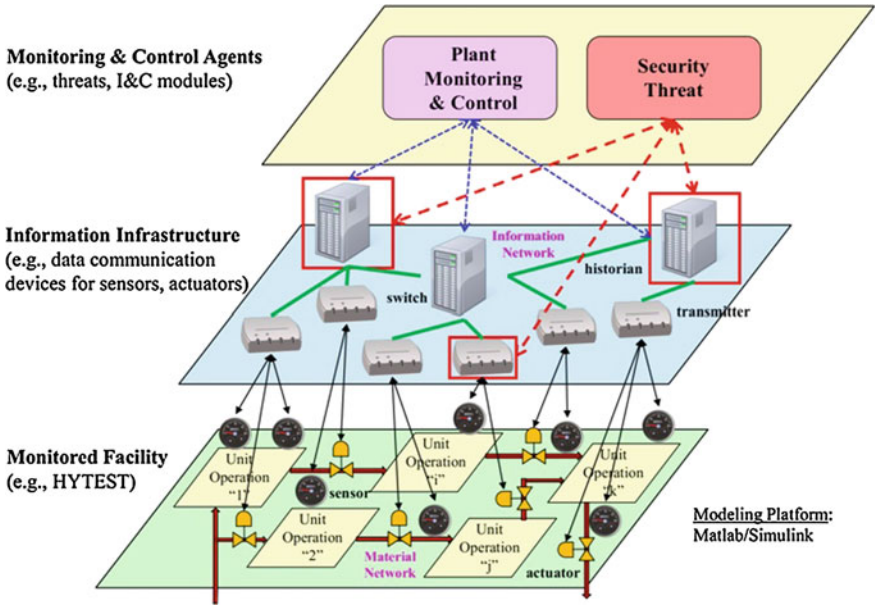

Fig. 1 Hybrid energy production facility

The fundamental assumption of the KSSM method is that the intelligent attacker is able to compromise any of the components in the information layer of the control system. The information layer is a communication layer which communicates physical process measurements to the process control layer, where they are presented to the operator. Figure 1 depicts an exemplary hybrid energy production system with highlighted physical, information and process control layers. In addition, it is assumed that the attacker will not be detected in the system as long as no transmitted measurement values are modified or blocked. It is important to emphasize here that the KSSM concept is intended not to detect anomalous process activity or whether the system functions within its normal operation envelope. Instead, the KSSM concept is designed to verify the system state information presented to the operator and reject system state falsification due to adversarial sensor measurement value corruption.

The main hypothesis of the KSSM concept is the idea that a small subset of sensor measurements, which are known to be secure (i.e. cannot be falsified in the physical layer), has the potential to significantly improve the observability of adversarial process manipulation due to cyber-attack. Furthermore, randomly selecting this small subset of known secure sensors can make more difficult for the attacker to identify which sensors measurements are being secured at particular time. Finally, it is assumed that there is only limited communication bandwidth available and the size of the selected KSSM sensor subset can be selected such that the real-time control of the system is not disrupted. We will describe in more detail these hypotheses in Sect. 3.2.

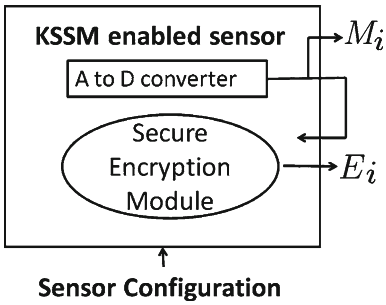

Fig. 2 Sample KSSM enabled sensor

In order to allow protection against an intelligent adversary, it must be possible to trust specific components of the system. In the KSSM system a cryptographic sensor module constitutes this trusted component as depicted in Fig. 2. The cryptographic sensor may be KSSM enabled with software or hardware as a mean to forward the plain-text measurement value M_i through a secure encryption module to produce a KSSM value E_i. If the particular sensor is part of the randomly selected subset of KSSM sensors, the encrypted measurement value E_i is sent to the control room after the plain-text measurement M_i.
The KSSM control module resides in the control room of the plant. The module is responsible for performing selection of the random subset of KSSM-enabled sensors. In addition, the control module also compares the received KSSM values with the plain-text measurements in order to detect falsification of the system state.

2.1 Attack Scenarios

While not required, we assume that all process sensors are KSSM hardware or software enabled, all encryption modules are secure from cyber attack, and the KSSM control room module, including the detection engine, are secure. Any other component in the system, including the entire information infrastructure layer may be assumed to be compromised by an attacker.
Figure 3 presents two scenarios. The system in scenario 1 has no KSSM sensors available. The adversary has compromised choke points in the communication network and is behaving as a man in the middle by preventing all valid sensor signals to the control room and replacing them with corrupted, signals C_i. This deception could be done, as Stuxnet partially demonstrated, through collection and then replay of sensor measurement data. The attacker is now able to manipulate the process as desired while the operators remain completely unaware. This level of attack may be undetectable without KSSM and will leave the operators completely blind to the actual system state.

KSSM sensors are available in scenario 2 and the attacker is choosing to corrupt only those signals which he knows are not providing encrypted values back to the

SCENARIO 1 **SCENARIO 2**

$$M_i \rightarrow C_i \qquad M_i \rightarrow C_i \vee M_i \vee (C_i \wedge E_i) \vee (M_i \wedge E_i) \qquad \text{for all } i = 1 \text{ to } N.$$

Fig. 3 M_i is the measurement from sensor i. C_i is the corrupted version of measurement M_i. E_i is the encrypted version of measurement M_i. $\vee$ and $\wedge$ are the logic disjunction and conjunction

detection engine. If the attacker corrupted the signals for which an encrypted version was sent at a later time then the corruption would be instantly recognized by the KSSM detector. Given that the encrypted signals are sent at some δ time after the unencrypted signal the attacker can only know which sensors will provide the encrypted signal by observing the network traffic for some period of time.
The attack in Fig. 4 consists on corrupting signals and blocking some of the encrypted versions. In fact attack would be easily detectable if the encrypted version of the measurement reached the detection engine and be compared against the corrupted version.
A way to make the above attack more difficult is to *periodically and unpredictably modify the subset of sensors providing encrypted values*. This ongoing and unpredictable selection of new sensors (and deselection of others) may be based on current system state, or communication network topology (for example not selecting KSSM sensors such that their encrypted measurements are all going through the same router).

Figure 5 schematically depicts the considered system state falsification scenarios and the counter-measures used by the KSSM system. The plain system state falsification is demonstrated in Fig. 5 (scenario 1). Here, the sensor measurements M_i are potentially corrupted by the attacker within the information layer. The falsified measurement values C_i reach the control operator. The basic idea of the KSSM system is depicted in Fig. 5 (Scenario 2), where a subset of the KSSM-enabled sensors is requested to report encrypted measurement values E_i to the control room. In this specific example, there will be a mismatch between values C_i and the decoded value of E_i. Further, an attacker aware of the KSSM protection system might attempt to deceive the system by blocking the encrypted values E_i from reaching the control room, as shown in Fig. 5 (Scenario 3). However, the KSSM system randomly modifies the subset of KSSM-enabled sensors, thus making it increasingly difficult for the attacker to design an attack with reliable detection delay. This is shown in Fig. 5 (Scenario 4), where the values C_1 and E_1 from the newly selected KSSM-enabled

SCENARIO 3

$$M_i \rightarrow C_i \vee M_i \vee (C_i \wedge E_i) \vee (M_i \wedge E_i) \vee \underset{\substack{\uparrow \\ \text{BLOCKED}}}{(C_i \wedge E_i)} \quad \text{for all } i = 1 \text{ to } N$$

Fig. 4 Attacker identifies which sensors are providing encrypted versions of measurements. During the attack only a few of the sensors are blocked

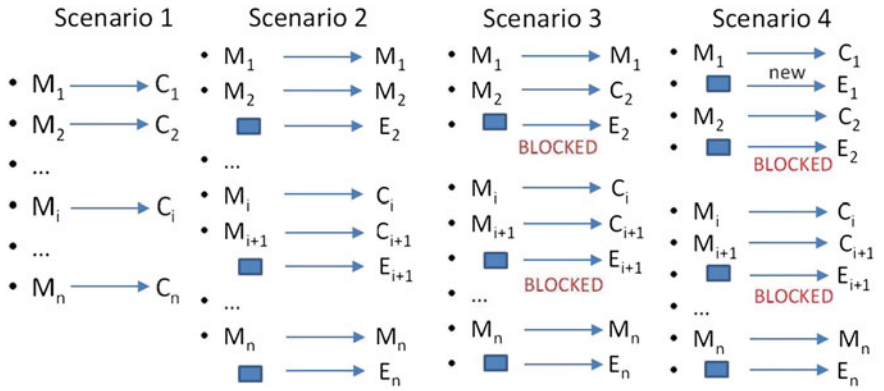

Fig. 5 Communication scenarios

sensor would produce a mismatch and indicate a presence of system state falsification.

3 Technical Objectives and Research Approach

Our objective is to investigate the value of KSSM for effective detection of unauthorized process manipulation and falsification of system state.

3.1 Targeted Facilities

As mentioned before a hypothetical hybrid energy plant is shown in Fig. 1. This figure represents a hybrid energy production facility with three abstract layers. The lowest layer is the physical process which consists of a set of production units each of which consists of reactors, tanks, gas flows, coolers, heaters, valves and other physical components. The information layer, in the middle, is responsible for communication. The sensors in the physical layer communicate with control devices and commands are sent to the edge controllers that drive actuator behavior. The highest layer is represented by the primary functions of plant control, and security threat monitoring and alarming. These highest level functions make use of real time data feeds from the physical plant up through the communication layer, and may also make use of information derived over time through initial monitoring of the system (e.g. passive network discovery).

We assume that the attacker can compromise any of the components in the information layer without being detected as long as the attacker does not modify the sensor signals being transmitted back to the controller and the control room. KSSM

is not designed to detect the system process exceeding its operational performance envelope, normal system monitoring is expected to detect that situation.

3.2 KSSM System Hypotheses

We created the following four hypotheses to stay focused on the core issues in conceptualizing, designing, and validating a prototype of a KSSM system.

1. **(H1)** A small set of sensor measurements, which are known to be secure, can significantly aid the operator and detection engine in more quickly and accurately identifying a cyber attack.
2. **(H2)** Some known secure measurements from randomly chosen sets of sensors providing data within selected time frames will harden the process against covert cyber attacks attempting to blind the operator and KSSM detection engine. Neither the enhanced operator effectiveness nor enhanced detection engine performance (gained from using a fixed set of known secure data) will be degraded by the changing and diverse sets of sensors selected for providing the known secure data.
3. **(H3)** It is possible to create a very low cost, limited bandwidth, and highly secure measurement capture and communication channel for transmitting k_i % ($0 < k_i < 100$) of a chosen sensor's physical measurements, end to end, from sensor to detection engine for analysis. The channel will involve adaptation of known cryptographic protocols to provide message and measurement integrity, and detection of replay attacks. Tradeoffs between cryptographic computational requirements at the sensor, power restrictions of a sensor, network bandwidth limitations, and the speed and accuracy of detection will be assessed in selecting specific cryptographic techniques for KSSM systems and establishing appropriate value for k_i.
4. **(H4)** Heuristics for selecting the set of sensors providing known secure sensor measurements can be developed which allow for the results of this research to be easily adapted for use in the design, implementation, and configuration of many diverse industrial control systems and infrastructures.

In a KSSM enabled infrastructure, the attacker will be unable to reliably falsify the process state to the control room operators.

3.3 KSSM Sensor

Figure 2 represents a KSSM hardware enabled sensor. The signal from the AD converter is tapped off and available to the secure encryption module. This module, at

some randomized time δ after the unencrypted measurement M_i is sent, forwards the encrypted version E_i of the measurement value to the KSSM detection module running on a control room computer. Whether or not the encrypted version of the plaintext measurement is sent depends on whether that particular sensor is currently selected by the KSSM control room module, and whether the secure encryption module selects it as one of the k_i of measurements for which a dual encrypted value will be expected. We note that these sensor functions may also be implemented in software and reside in the sensor or closest computational edge point. The KSSM detection algorithm in the control room, which must also be trusted, will compare the two versions of the measured value, unencrypted and encrypted, and trigger an alarm if there is any difference. For the exposition of the idea in this chapter, we are making simplifying assumptions related to reliable transport of measurements, both plaintext and ciphertext, and to the reliability of the sensor and encryption hardware.

3.4 KSSM Control Room Module

The KSSM module residing in the control room is represented in Fig. 6. It is responsible for modifying the subset of KSSM-enabled sensors which perform encryption, and is also responsible for detecting attacks. Many functions are needed to provide these capabilities and we will very briefly describe only the highest level functions.

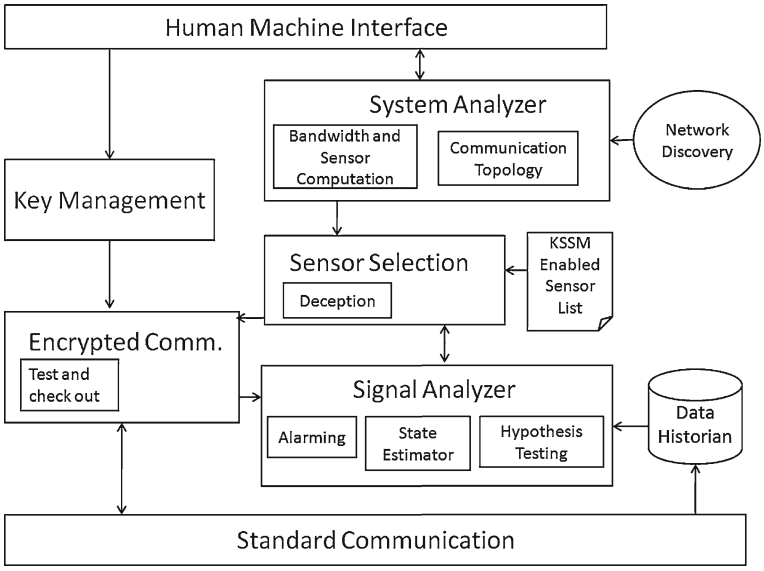

Fig. 6 Block diagram of KSSM module residing in the control room

The *system analyzer* receives input from network discovery tools which can both reside on the system and operate in real time, or can be one time only devices used during a phase such as system acceptance testing. It develops simplified models of the communication network to aid the sensor selection function in choosing smart subsets of sensors.
The *signal analyzer* is responsible for analyzing the sensor measurements that are provided to the control room, and alarming when appropriate. If encrypted and associated unencrypted values do not match then an alarm will be set; if some number of requested encrypted values do not arrive in a timely fashion, and are distributed over a variety of communication paths then it may be appropriate to raise an alarm based on probabilistic assessment of likely communication and sensor failures.
The *sensor selection algorithm* will incorporate what is known about the communication topology and the failure rates of all components within the system. The failure rates may be based on empirical data or models built into the algorithm. Further some understanding of the limits on computation cycles available, sensor power restrictions, and limitations in communication bandwidth will be incorporated to aid, not only the selection of a new subset of sensors for KSSM but also the selection for each chosen sensor of the k_i of measurements that will be encrypted and forwarded. Sensor selection and the percent of measurements for which dual encrypted values are required may also be made selectable by the operators so that they have control in limiting the sensor processor cycles, sensor power consumption, and communication bandwidth utilized by KSSM.
The *cryptographic functions* will be adopted from currently well understood cryptographic components and systems. The KSSM-enabled sensor list is needed so that sensor selection can accommodate systems that are slowly being upgraded with KSSM-enabled sensors. And the KSSM user interface will be separate from all other devices in the control room in order to provide as much hardening against attack as possible.

4 Known Secure Sensor Measurement System Simulation

This section describes the design and simulation of the KSSM-equipped control systems. First the overall architecture is presented. Next its major components of Sensor Selector and Signal Analyzer are described in more detail.

4.1 KSSM System Architecture

The overall KSSM system architecture is depicted in Fig. 7. The system is composed of two major parts, the KSSM control module and the communication network which connects the control module with those sensors that are KSSM-enabled. The KSSM control module is composed of two main components, the Signal Analyzer and

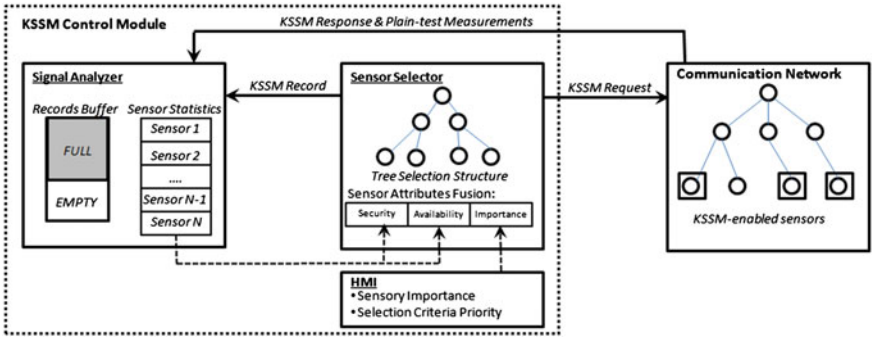

Fig. 7 Architecture of the KSSM system

the Sensor Selector. All components monitor network traffic in the control system and communicate among each other to perform effective system state falsification detection while minimizing the impact on the system's communication bandwidth.

The Sensor Selector component is responsible for selecting a subset of KSSM-enabled sensors every time iteration. The sensor selection is performed using a tree-like sensor selection data structure, which resembles the known network topology. The Sensor Selector uses several criteria, including subjective human input to calculate the selection weight of each sensor. A randomization algorithm is then applied to ensure representative sensor selection from the communication network. Every time a subset of sensors is selected by the Sensor Selector a KSSM request is sent to the sensors and a KSSM record about the selection is stored in the Signal Analyzer. The Signal Analyzer is responsible for monitoring both the plain-text and the KSSM encrypted network messages. Every time a KSSM record about sensor selection is received from the Sensor Selector, the Signal Analyzer stores the record in a record buffer. Upon receiving the previously requested KSSM message from the network, the KSSM value is paired with its plain-text value stored in the record buffer and their values are compared. The Signal Analyzer also keeps track of important network traffic statistics such as sensor availability and response latency, which are used for adjusting the sensor selection process.

4.2 KSSM Sensor Selector

The main task of the Sensor Selector is to perform randomized sensor selection every time iteration. To achieve this, the Sensor Selector contains an approximate model of the network topology in a form of a tree data structure. The root of the tree corresponds to the main communication node of the control system network. Branches connect the root node to possibly multiple levels of nodes. Each node corresponds to a sub-network in the real network system. Finally, leafs of the tree structure correspond to

individual KSSM-enabled sensors. It should be noted that it is not required for the tree structure to exactly match the real communication network topology. Rather, the branches of the tree should correspond to logical units in the control system network, in order to achieve evenly distributed sensor selection. The process of sensor selection is performed by randomly descending from the root of the tree to particular leaf. All branches in the selection tree emanating from a particular node are assigned a specific selection probability, which guides the random descending process. This method is repeated until the new subset of KSSM enabled sensors has been selected. The branch selection probabilities are updated after selection of each sensor so that more probability is distributed to the branches that were not assigned. The pseudo-code of this randomized sensor selection algorithm can be summarized as follows:

1. Initialize the sensor selection probabilities p_{ij} of each branch in the selection tree.
2. Repeat for all k KSSM sensors.
   a. Set current node n_i as root.
   b. Repeat, until current node n_i is a leaf.
      i. Randomly select j^{th} branch of current node ni based on branch selection probabilities p_{ij}.
      ii. If there exist unselected leafs in the sub tree connected to the j^{th} branch descend to the j^{th} children of current node n_i.
   c. Return the index of the sensor in the selected leaf.
   d. Repeat until current node n_i is a root.
      i. For all siblings of current node n_i compute the new branch selection probability from their parent as:

$$p_{kj} = \begin{cases} p_{kj}(1-\alpha), & \text{if } k = i \\ p_{kj} + \frac{\alpha p_{ij}}{k-1}, & \text{if } k \neq i \end{cases} \quad (1)$$

      ii. Ascent to the parent of node n_i.

Coefficient α controls the spatial diversification of the selected sensors. Values close to 1 will result in large spatial diversification (e.g. sensors sampled in different areas of the network), while values closer to 0 will result in selected sensors being more likely to be close to each other (e.g. in the same sub-network). Parameter k denotes the cardinality of the selected KSSM sensor subset. This process of KSSM enabled sensor selection and selection weight updates is depicted in Fig. 8. Due to the re-distribution of branch selection weights, the subset of sensors is more likely to be distributed throughout the network. Hence, KSSM and plain text message loss rate due to random component failures in parts of the communication system can be reduced.

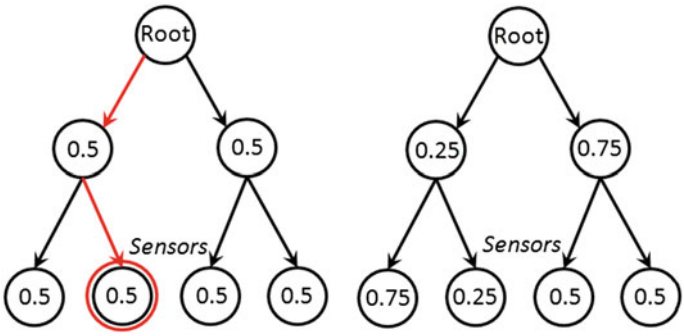

Fig. 8 Architecture of the KSSM system

After the subset of KSSM enabled sensors has been specified the Sensor Selector re-computes the initial branch selection probabilities in the selection tree to reflect the most current behavior of the communication system. These recomputed branch selection probabilities are used to initialize the tree parameters in Step 1. This process for computing the initial branch selection probabilities is composed of three parts: (1) sensor selection weight calculation, (2) bottom-up selection weight propagation, and (3) top-down selection probabilities normalization.

The sensor selection weight is calculated for each KSSM-enabled sensor based on a weighted average of three parameters: availability, security and importance. The availability can be computed as the inverse value of the averaged time interval of obtaining the requested KSSM value from the particular sensor. When the sensor response time increases, its availability is decreased and the sensor will be selected less often to ease the work load of the particular sensor and its part of the network.

The security is computed as the averaged time interval between receiving two mismatching KSSM-values and plain text values. Because random noise might corrupt the KSSM messages, single mismatch should not immediately raise an alarm. However, when the frequency of mismatched messages is significantly increased the security is increased, which results in sensor being selected more often to quickly converge to final detection. Here, a significant increase is considered to be an increase above the normal frequency of mismatched measurement values due to ordinary communication noise.

Finally, the importance attributed to a sensor is a subjective value provided by the operator, which can help to fine-tune the selection algorithm (e.g. some sensors might be more important for the control and thus should be sampled more often). In addition, the operator can specify the weighting coefficients for the weighted average of these attributes.

The bottom-up selection weight propagation proceeds in a recursive manner and its purpose is to propagate the sensor selection weights up the tree. The algorithm reads the selection weight from all children into their common parent, the weights

are summed and recursively propagated to the higher level until the root node is reached.
In the final stage, the selection weights need to be converted into branch selection probabilities. This is achieved by descending from the tree root to individual leafs and normalizing the selection weights for all branches emanating from each node. The normalization procedure ensures that all branch selection probabilities sum up to 1 for each node.

4.3 KSSM Signal Analyzer

The main task of the Signal Analyzer is to monitor the network traffic and detect potential falsification of system state. Every time a KSSM request is sent to a particular sensor a record about this is stored in the record buffer in the Signal Analyzer. Upon receiving the KSSM measurement value, the corresponding plain-text measurement is looked up in the record buffer. The KSSM measurement is decrypted and compared to the plain-text value. A measurement mismatch can be used to indicate a potential presence of an intelligent adversary in the information layer of the system. The intelligent adversary who is aware of the KSSM system might attempt to avoid detection by preventing the KSSM values from reaching the Signal Analyzer. For this reason, the record buffer contains an upper limit on the number of active KSSM records. When a KSSM message is blocked its plain-text counterpart will not be removed from the record buffer and the capacity of the buffer will be decreased. When this capacity reaches the specified threshold, an indication of potential attempt to falsify the system can be reported.
The Signal Analyzer also gathers important network traffic attributes, which are used to adapt the KSSM system to the specifics of the current network traffic. First, the time interval of requesting and receiving a KSSM value is computed for each sensor. This information is used to calculate the availability of individual KSSM-enabled sensors. Next, the time interval between obtaining two mismatched plain-text and KSSM values for each sensor is being monitored. This information is used to calculate the security of individual sensors and used for sensor selection. Finally, the Signal Analyzer stores the response time of obtaining the plain-text measurements, which can be used to monitor and adjust the appropriate size of the requested KSSM sensor subset so that the response of the control system is not affected. This adaptive mechanism is explained below.
The Signal Analyzer monitors the maximum response time of any plain-text sensor and compares that to the requested allowed response time. For example, if the sensor values should be reported to the control room once every second then the maximum allowed response time can be set to 0.8 s to create a safety buffer. The difference between maximum and the allowed response time creates a feedback signal that could be used to adjust the number of sampled KSSM sensors so that the real-time system response is not affected. When the maximum response time is below the allowed threshold for a certain period of time, the number k of sampled KSSM

sensors is increased by one. Similarly, when the maximum response time is greater then the allowed threshold for certain amount of time, the number k of sampled KSSM sensors is decreased in order to preserve the real-time response of the system.

5 Sensor Selection: Experimental Results

This section first describes the implemented virtual communication network used as an experimental test-bed. Next, a set of testing scenarios is used to demonstrate the performance of the proposed KSSM system.

5.1 Experimental Test-Bed

In order to validate the performance of the designed KSSM system a virtual communication network was implemented. The network simulator models packet-based traffic in control system communication networks. The network is composed of communication nodes and sensor nodes. The communication nodes are equipped with packet buffers and routing tables. The packet buffer dispatches packets on first-in first-out basis. The sensor nodes can generate the plain-text measurement value as well as its encrypted version upon request. The network simulator can simulate various deterministic as well as stochastic properties of the network. For example, the desired throughput can be set for individual network nodes as well as stochastic packet loss rates or packet corruption rates. The KSSM Control module is connected to the communication network interface, where KSSM requests can be passed into the network and plain-text and KSSM messages can be received. For the purpose of experimental testing a simple control system communication network was constructed. The network gathers measurements from 9 sensors, which are grouped into 3 sub-networks as depicted in Fig. 9.

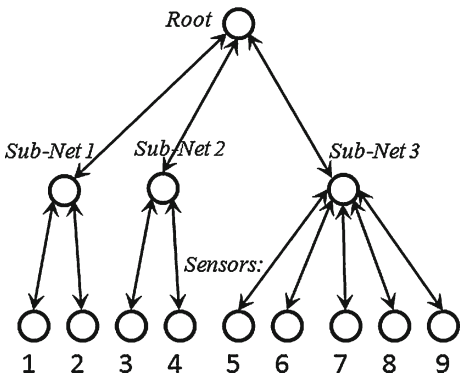

Fig. 9 Testing network topology

5.2 Sensor Selection

The purpose of the first testing scenario was to demonstrate the automatic adaptation of the sensor selection algorithm to reflect the current behavior of the observed network traffic. In this scenario, the control system is run for 10,000 s and the sensor data is gathered once every second. In addition, $k = 2$ known secure sensor values are requested every second. The communication network is initialized with uniformly distributed time delay and packet loss and corruption rates throughout the entire network. Also, all of the selection criteria for individual sensors are weighted equally. Three events are used to simulate various changes of the environment to demonstrate the adaptation mechanism of the Sensor Selector.

- Event 1: At time t = 2500 s the network traffic in the larger sub-network 3 becomes congested, which is implemented as decreased throughput of particular communication node. Hence, the availability of sensors 5–9 is decreased.
- Event 2: At time t = 5000 s a possible cyber-attack is simulated on sub-network 1. This attack is implemented as an increased packet corruption rate for the associated communication node leading to increased number of mismatched plain-text and KSSM messages from sensors 1 and 2.
- Event 3: At time t = 7, 500 s the operator decides to adjust the sensor selection mechanism via the HMI by assigning weight 1.0 to the importance attribute and decreasing the weight of the security and availability attributes to 0.1. In addition, the operator subjectively increases importance of sensor 4 to its maximum value of 1.0.

Event 1 affects the availability of sensors 5–9. After the time delay for messages from sub-network 3 was increased, the response times of the KSSM messages from sensors 5–9 were increased. This resulted in decreased availability of sensors 5–9 as shown in Fig. 10. Event 2 affects the security of sensors 1 and 2. The increased probability of obtaining an incorrect KSSM message from sensors 1 and 2 causes the time interval of receiving two mismatching plain-text and KSSM messages to decrease. Hence, the security of these sensors is increased, which can be observed in Fig. 11. Figure 12 shows the evolution of the sensor selection weight for individual

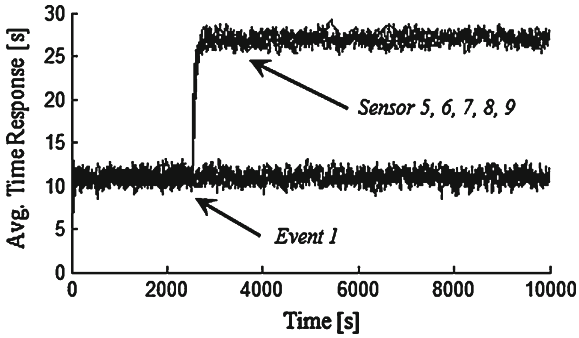

Fig. 10 Average time response for various KSSM sensors

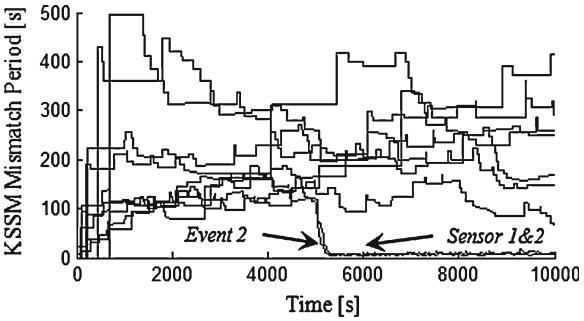

Fig. 11 KSSM values mismatch period

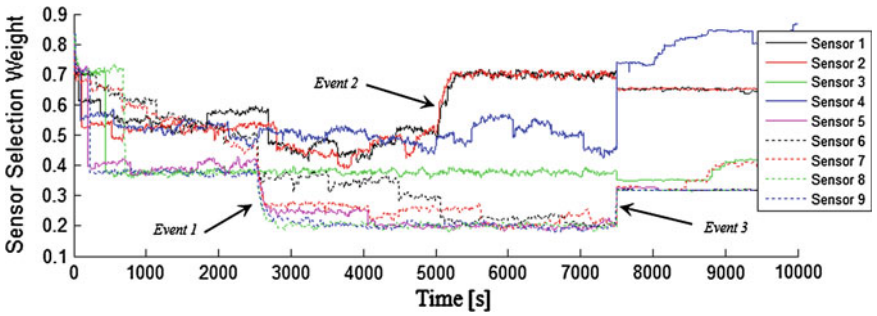

Fig. 12 Selection weight for different sensors during the test scenario

sensors. It is apparent how the sensor selection weights are converging to a uniform distribution during the first 2,500 s of the simulation. The diverse selection weights at the start of the simulation are due to the stochastic sampling process, which must be first averaged over certain amount of time to obtain good initial results. Next, it is apparent that the decreased availability of sensors 5–9 when event 1 occurs leads to their lower selection weight. Furthermore it can also be seen that the increased security of compromised sensors 1 and 2 when event 2 occurs leads to their increased selection weight. Finally, Event 3 at time 7,500 s can be observed when the operator overrides the selection criteria importance and modifies the selection weight, which increases the weight of sensor 4 due to its higher importance. To verify the influence of the sensor selection weight on the KSSM sensor sampling process, Fig. 13 shows histograms of sensor selection for the four quarters of the simulation. It can again be observed that the decreased value of the availability parameter leads to less frequent selection of sensors 5–9 in Fig. 13b and the increased security of sensors 1 and 2 leads to their more frequent selection in Fig. 13c. Finally, the higher importance of sensor 4 results in its more frequent sampling together with sensors 1 and 2 that were likely compromised by an attacker, as shown in Fig. 13d. In summary, Fig. 13 demonstrates that the KSSM system adjusts the sensor selection algorithm to obtain more samples

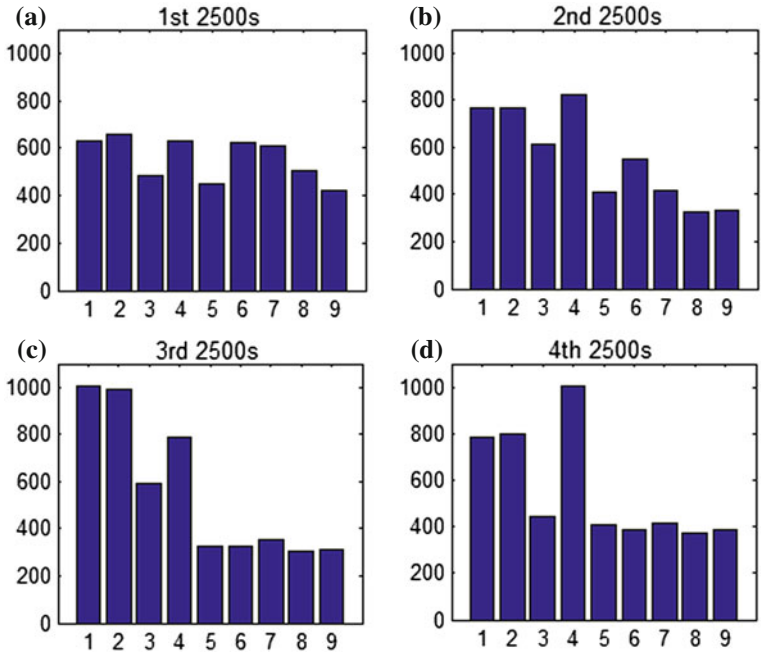

Fig. 13 Sensor selection histograms for different intervals of the simulation

from likely compromised sensors and to obtain less samples from congested parts of the communication network.

5.3 Variable Network bandwidth

The following test scenario was designed to demonstrate the automatic update of the number k of sampled KSSM messages. The essential property of the KSSM system is that it should use the available communication bandwidth in the control system network without compromising its real-time response. In this scenario, the identical communication network as shown in Fig. 9 was used. The network was simulated for 4,000 s and the sensor measurements have been reported once every second. In order to achieve the requested real-time response of obtaining sensor measurements once every second, maximum desired response for plain-text measurement was set to 0.8 s. For the initial 1000 s, the network was simulated with low average time-delay for individual network nodes (0.05 s average latency of network node per packet). At time 1000 s the average time delay of the larger sub-net 3 was increased to 0.075 s. Next, at time 2000 s the average time delay of sub-net 3 was increased to 0.1 s. And finally at time 3000 s the average time delay of sub-net 3 was increased to

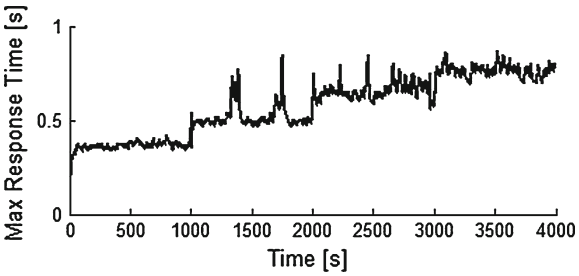

Fig. 14 Maximum response time of plain-text measurements

0.125 s. Note that the actual time delay for a specific packet was computed using a uniform distribution with standard deviation of 0.02 s centered at the average time delay value. Figures 14 and 15 demonstrate the behavior of the system. First, Fig. 14 depicts the maximum observed response time of the plain-text measurements. It is apparent how this maximum response time increases at times 1000, 2000, and 3000 s. Next, Fig. 15 shows the number k of selected KSSM sensors. The algorithm starts with $k = 0$ KSSM sensors and first observes the maximum response time of the plain-text measurements. When this maximum response time is found to be below the desired threshold of 0.8 s, the number k of selected KSSM messages is incrementally increased up to the maximum value of all 9 sensors sending encrypted messages. The first increase in time-delay at time 1000 s caused several plain-text messages to be delivered later than the required real-time response and the system quickly lowers the value of k in order not to disrupt the real-time response. Consequently, the system attempts to sequentially increase the number of sampled KSSM values, while monitoring the real-time performance. Finally, in the last part of the simulation the system stabilizes and samples mostly a single KSSM value per iteration. Hence, it can be observed that the KSSM system attempts to provide the maximum level of cyber-state awareness given the available communication bandwidth.

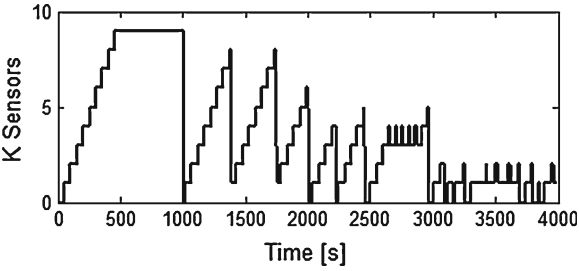

Fig. 15 Number of selected KSSM values

6 Future Work

As previously stated one of the most important features of the presented KSSM system is that it can be incrementally integrated with existing control systems. This incremental integration is allowed by utilizing the existing physical measurements and by using the available communication network bandwidth. For larger scale control systems with high number of physical sensors the incremental integration with the KSSM system might have to be performed in several stages. In each stage a small subset of physical sensors would be enhanced with an encryption module. The newly enhanced sensor would then be added to the list of KSSM-enabled sensors and it could be consequently used to for system state falsification detection.

Once all available physical sensors are KSSM-enabled the KSSM system can provide the maximum level of cyber-security given the communication network resources available. However, in the earlier stages of the KSSM system implementation where not all physical sensors are KSSM-enabled, it is important to prioritize which sensors should be made KSSM-enabled in the current stage so that the current level of provided cyber-security is maximized. This prioritization constitutes a complex multi-criteria decision making task because multiple potentially conflicting constraints such as economical, time or other subjective constraints might be simultaneously present. Optimal prioritizing and staging the implementation process is a crucial component of applying the KSSM concept to real world systems.

Our future work will be to further research and apply the KSSM concept to a variety of explicit real world systems on both a macro and micro scale. On the macro scale, we will apply multi-criteria decision making to the question of optimal placement of a limited number of PMUs in portions of the U.S. power grid. This will be evaluated with the understanding that the placement of PMUs will be staged in over many years. On the micro scale we will research and evaluate building KSSM concepts in at a local level, such as at significant individual substations. Within each substation, the intent will be to build in local KSSM agents within sensors and actuators so that they may autonomically manage a localized KSSM process for substation state awareness. The localized state awareness may then be transmitted back to the control room for regional level state awareness or used for localized response to detected cyber attacks. These KSSM agents will of course be applied with the understanding that they must not ever interfere with necessary communication and operations of the substation in the absence of cyber attack.

Acknowledgments This work was partially supported by the U.S. Department of Energy under DOE Idaho Operations Office Contract DEAC07-05ID14517, performed as part of the Known Secure Sensor Measurements and Experimental Security projects at Idaho National Laboratory, the Los Alamos National Laboratory LDRD project Optimization and Control Theory for Smart Grids, the Center for Nonlinear studies at Los Alamos National Laboratory.

References

American Gas Association (2005a) Cryptographic protection of SCADA communications, part 1: background, policies and test plan, draft 5, April 2005

American Gas Association (2005b) Cryptographic protection of SCADA communications, part 2: retrofit link encryption for asynchronous serial communications, April 2005

Cheung S, Dutertre B, Fong M, Lindqvist U, Skinner K, Valdes A (2007) Using model-based intrusion detection for SCADA networks. In: Proceedings of the SCADA security scientific symposium, 2007

Giani A, Bitar E, Garcia M, McQueen M, Khargonekar P, Poolla K (2011) Smart grid data integrity attacks: characterizations and countermeasures. In: Proceedings of the smart grid communication, 2011

Linda O, Manic M, McJunkin TR (2011a) Anomaly detection for resilient control systems using fuzzy-neural data fusion engine. In: Proceedings of the IEEE symposium on resilient control systems, August 2011

Linda O, Manic M, Vollmer T, Wright J (2011b) Fuzzy logic based anomaly detection for embedded network security cyber sensor. In: Proceedings of the IEEE symposium on computational intelligence, pp 202–209, April 2011

Linda O, Vollmer T, Manic M (2009) Neural network based intrusion detection system for critical infrastructure. In: Proceedings of the IJCNN 2009, June 2009

McQueen M (2009) Deception used for cyber defense of control systems. In: Proceedings of the 2nd conference on human system interactions, 2009

McQueen M, Giani A (2011) Known secure sensor measurements for critical infrastructure systems: detecting falsification of systems state. In: Proceedings of the SERENE, 2011

Rieger CG, Gertman DI, McQueen MA (2009) Resilient control systems: next generation design research. In: Proceedings of the 2nd IEEE conference on human system interactions, Catania, Italy, May 2009, pp 632–636

Schneier B, Ferguson N, Kohno T (2010) Cryptography Engineering, Chaps. 3–7. Wiley and Sons, New York, NY

Stamp M (2011) Information security, Chaps. 3–5, and 9, 2nd edn. Wiley and Sons, Hoboken, NJ

Tsang C (2005) Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction. In: Proceedings of the 2005 IEEE international conference on industrial technology (ICIT 2005)

Wei D (2010) An integrated security system of protecting smart grid against cyber attacks. In: Proceedings of the innovative smart grid technologies (ISGT), 2010

Yang D, Usynin A, Hines JW (2006) Anomaly-based intrusion detection for SCADA systems. In: Proceedings of the 5th international topical meeting on nuclear plant instrumentation, control and human machine interface technologies (NPIC and HMIT 05), Albuquerque, NM, 12–16 November 2006

Zhong S, Khoshgoftaar T, Seliya N (2007) Clustering-based network intrusion detection. Int J Reliab Qual Saf 14(2):169–187

Data Diodes in Support of Trustworthy Cyber Infrastructure and Net-Centric Cyber Decision Support

H. Okhravi, F. T. Sheldon and J. Haines

Abstract Data diodes provide protection of critical cyber assets by the means of physically enforcing traffic direction on the network. In order to deploy data diodes effectively, it is imperative to understand the protection they provide, the protection they do not provide, their limitations, and their place in the larger security infrastructure. In this work, we study data diodes, their functionalities and limitations. We then propose two critical infrastructure systems that can benefit from the additional protection offered by data diodes: process control networks and net-centric cyber decision support systems. We review the security requirements of these systems, describe the architectures, and study the trade-offs. Finally, the architectures are evaluated against different attack patterns.

Keywords Data diodes · Trusted process control networks · Industrial control systems · Cyber decision support systems · Net-centric systems

1 Introduction

Data diodes have been proposed as network devices which strictly limit the flow of information in secure networks. They enforce one-way traffic by physically restricting the back flow of information. Data diodes have gained popularity in military

H. Okhravi and J. Haines
Massachusetts Institute of Technology, Lincoln Laboratory,
244 Wood St, Lexington, MA 02420, US
e-mail: hamed.okhravi@ll.mit.edu

J. Haines
e-mail: jhaines@ll.mit.edu

F. T. Sheldon (✉)
Oak Ridge National Laboratory (ORNL), One Bethel Valley Rd, Oak Ridge,
TN, US
e-mail: sheldonft@ornl.gov

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*,
Energy Systems, DOI: 10.1007/978-3-642-38134-8_10,

and avionics industries (Menoher and Mraz 2007; Roach 2007) where they separate highly valuable networks (e.g. a classified network) from less important networks (e.g. an unclassified network). Using data diodes one can ensure that information can flow from one side to another, but not the reverse. This property can be used, for instance, to ensure that classified information cannot leak to the unclassified network.

In this paper, we first provide an overview of data diodes, the protection they can provide, the protection they cannot provide, and their limitations. We then propose two critical infrastructure applications that can benefit from the additional protection offered by data diodes: process control networks and net-centric cyber decision support systems. In the process control applications, data diodes will be used to protect the critical components of the system (controllers) from breaches in the enterprise networks, improving their integrity and availability. In cyber decision support systems, data diodes ensure the confidentiality of the intelligence and integrity of the decisions. We evaluate the proposed architectures against different attack patterns.

The rest of the paper is organized as follows. Section 2 provides an overview of data diodes and their functionalities and limitations. The application of data diodes in trusted process control networks and net-centric cyber decision support systems are discussed in Sects. 3 and 4. Section 5 provides the evaluation of the proposed architectures. We review the related work in Sect. 6 before concluding the paper in Sect. 7.

2 Data Diodes

Data diodes provide a physical mechanism for enforcing strict unidirectional communication between two networks. They are often implemented by removing transmitting component from one side and receiving component from another side of a bidirectional communication system (e.g. a fiber optic system with TX capability in only one side and RX capability in the other side.) Data diodes can only send information from one network (a.k.a the "low" network) to another network (a.k.a the "high" network). The high network often contains data with a higher classification level than the low network. Figure 1 illustrates two networks connected by a data diode.

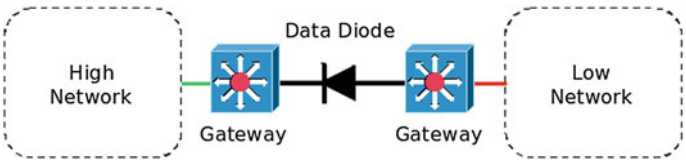

Fig. 1 Two networks connected by a data diode

2.1 Protection Provided

Data diodes can provide strong confidentiality from the high network to the low network; i.e. provided that the unidirectional connection is the only communication link between these two networks, information can flow from low to high, but there is no back-flow of data. In a dual fashion, data diodes can provide strong integrity from the low network to the high network; i.e. a malicious component in the high network cannot corrupt data or perform network-based attacks on the low network (availability). In this way, we can provide containment within the high network while at the same time providing better availability within the low network.

2.2 Protection Not Provided

It is sometimes claimed that data diodes protect the high network against cyber attacks. This, in fact, is not correct. Many cyber exploits do not require a session or bidirectional communication. Often fast propagating worms or malware need just one packet of data to infect a machine. Self expanding malware or quine programs (Hofstadter 1979) even limits the number of bytes required in the packet (Rieback et al. 2006).

Moreover, in industrial control systems, the process control network is the critical component of the system for which availability and integrity are critically important properties. If the process control network (PCN) is connected to the "high" side, the data diode does not protect it against breaches from the low network. Information may not be exfiltrated from the PCN but normally that does not matter in the case of disruptive quine programs that can affect both integrity and availability.

2.3 Limitations

A major limitation of the data diode is that it does not work with the standard TCP/IP protocols. A data diode requires proprietary unidirectional protocols that do not require acknowledgments. On both sides of a data diode, gateways translate unidirectional protocols to standard bidirectional protocols to connect the diode to the rest of the network (Waterfall's Unidirectional Security Gateways 2010). However, more high-end products (Interactive Link Data Diode Device 2010) also accept TCP or UDP packets as input. Data diodes can be used to enhance security, but they are by no means even a nearly complete solution. They have to be placed carefully in conjunction with other defensive mechanisms.

2.4 Implementation

Data diodes are often implemented using serial links (RS-232) or optical fiber. In the serial link implementation, one of the two data cables (from high to low) is removed. In optical data diodes, the transmitter of the high network and the receiver of the low network are removed.

A major disadvantage of the RS-232 implementation is that in addition to data lines, there are control lines defined in the standard along which data can potentially flow back to the low network. Hence, optical fiber is the preferred implementation for data diodes.

3 Trusted Process Control Networks

3.1 Overview of PCNs and Security Challenges

Figure 2 illustrates a typical process control network (PCN) architecture with paired firewall. In this architecture, the PCN contains the low level control devices such as programmable logic controllers (PLCs), remote terminal units (RTUs), master terminal unit (MTU), and the operator console. The enterprise network (EN) often contains the workstations and high level management consoles. The data historian sits in the demilitarized zone (DMZ) of the firewalls and acts as an intermediary between the PCN and EN. In fact, to protect the PCN from attacks and breaches going through the EN, status data is only collected from the historian and not directly from the PCN.

Protecting PCNs from a variety of attack vectors often faces many challenges. Firewall configuration errors may result in unwanted traffic going to the PCN or legitimate traffic being dropped. In fact a study by Wool (2004) shows that 80 % of firewall rule sets allow any service on inbound traffic and insecure access to firewalls. Moreover, a firewall may be bypassed by an attacker using encrypted tunnels (e.g., VPN) or unsecured out-of bounds communication (e.g., dial-up maintenance connection). Vulnerable end devices also pose a threat to the security of PCNs. Software/configuration bugs in the control devices may be exploited by an attacker to gain illegitimate access to the system or to change the configuration of critical components. Unsecured physical access to any part of the network (e.g., unsecured Ethernet ports) may also result in a benign or malicious damage to the PCNs. In addition, untrusted (rogue) devices or users may enter the network and breach its security. Finally, all of the above mentioned vectors may introduce malware (worms and viruses) to the critical systems.

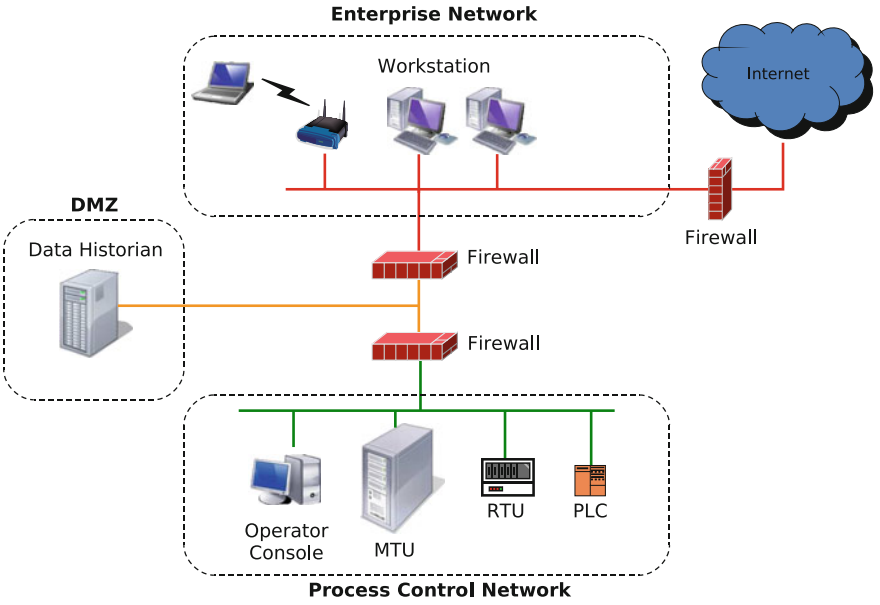

Fig. 2 A typical paired-firewall industrial control system

3.2 Trusted Process Control Networks (TPCN) with Data Diodes

The TPCN architecture (Okhravi and Nicol 2009) deploys trusted network (TN) (Network Admission Control 2005) technology to establish trust in devices for control systems. It uses state information from the hardware and software in devices for admission and access control decisions. When a device first joins the network, its hardware and software are checked; based on these checks, the appropriate access control rules are applied dynamically to the user, device and traffic. The TPCN architecture uses existing standards, protocols, and hardware devices to extend the concept of "trust" to the entire network architecture. A TPCN has the following components:

- Client device: Every client device must be evaluated prior to admission to a TPCN.
- Network Access Device (NAD): All connectivity to a TPCN is implemented via a NAD, which enforces authentication, authorization and access control policy. NAD functionality may exist in devices such as switches, routers, VPN concentrators and wireless access points.
- Authentication, Authorization, and Access Control (AAA) Server: maintains the policy and provides rules to NADs based on the results of authentication and posture validation.
- Posture Validation Servers (PVSs): evaluate the compliance of a client before it can join a TPCN. A PVS is typically a specialization for one client attribute (e.g., operating system version and patch or virus signature release).

- Posture Remediation Servers (PRSs): provide remediation options to a client device in the case of non-compliance.
- Directory Server (DS): authenticates client devices based on their identities or roles.
- Other Servers: These include trusted versions of Audit, DNS, DHCP and VPN servers.

The TPCN architecture is presented in Fig. 3. A client device intending to join the network communicates its request to the NAD. The NAD establishes the client device's identity using Extensible Authentication Protocol (EAP) over the 802.1x protocol and sends the results to the AAA server using the Remote Authentication Dial In User Service (RADIUS) protocol. The AAA server returns a list of posture validation requirements and the addresses of the appropriate PVSs. The client then validates its posture with each of the PVSs. If the client is in compliance, the results are sent to the AAA server using the Host Credential Authorization Protocol (HCAP) protocol. On the other hand, if the client lacks one or more requirements, the appropriate posture remediation servers suggest remediation actions to the client. The directory server determines the client's group or role. Given all the results from the PVSs and the directory server, the AAA server determines the set of rules that apply to the client's access and traffic and sends them to the NAD for enforcement.

From this point on, the client is permitted to communicate via the NAD and all its activities are monitored for policy compliance. The policy held by the AAA server is in the form of an authentication requirement and a list of posture validation requirements.

When a client device joins the network, a NAD communicates with an AAA server on behalf of the device. The AAA server authenticates the device and provides rules based on the device's security postures to the NAD. From this point on, the NAD enforces the policy on all ingress and egress traffic to/from the device. For example, an RTU with valid firmware is allowed to communicate with the historian; all other traffic is blocked. Okhravi and Nicol (2009) provide two examples to further clarify the workings of a TPCN. They also describe methods to enhance availability of TPCNs and limit the number of configuration errors.

The TPCN addresses many of the security challenges by providing defense-in-depth and extending trust to the process control devices (Okhravi and Nicol 2008). TPCNs build a security infrastructure for mission critical process control systems. Data diodes can be used to enhance TPCN protection by strictly limiting traffic at some sensitive points.

An important component of the TPCN network that can benefit from data diodes and tolerate their limitations is the data historian. The firewalls are often configured to drop any traffic going from the data historian to the PCN. If a data diode is placed between the historian and the PCN, the critical control devices can still push their status data to the DMZ while no traffic can flow back. Another diode may also be placed between the DMZ and EN to protect the integrity of the historian. Note that in both cases the "high" end of the diode is connected to the less critical components (see Fig. 1). This protects the PCN against attacks from EN or DMZ, granting integrity

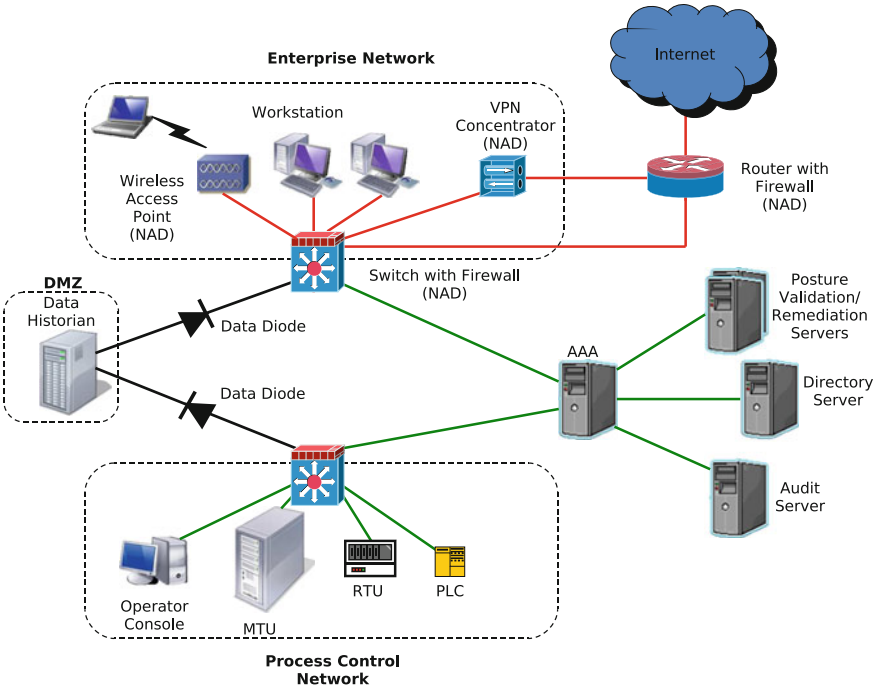

Fig. 3 A TPCN with data diodes

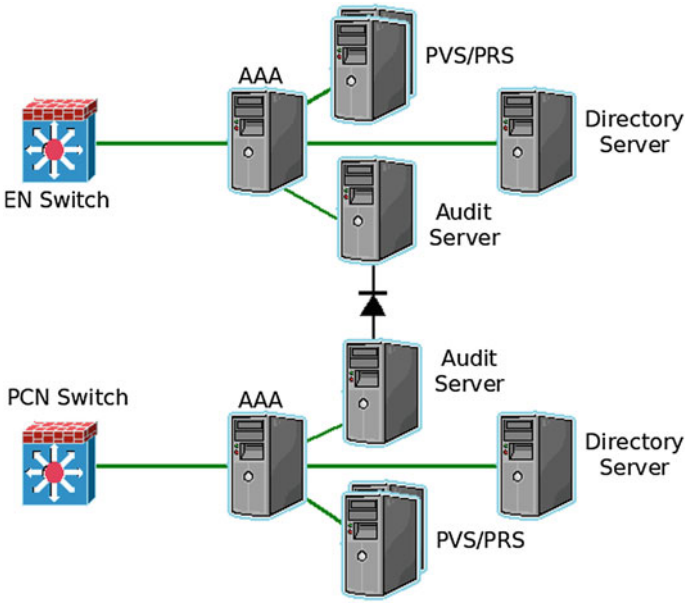

Fig. 4 Separate authentication/unified auditing model

and availability. The confidentiality of the data sent to historian is arguably less important than protecting the PCN.

3.3 Separate Authentication and Unified Auditing

In the architecture illustrated in Fig. 3, we assume that the servers (AAA, PVS, PRS, DS, and audit server) are secure and they cannot be used in a stepping-stone attack to penetrate the PCN. To achieve higher security and relax this assumption, the TPCN servers must be replicated for EN and PCN (see Fig. 4).

In this architecture, the servers on the EN side perform the authentication and posture validation for EN while a different set of servers perform these functionalities for PCN. The separate authentication scheme has two downsides: (1) it imposes additional hardware cost to the system and (2) the system no longer has a unified audit log. The extra cost can be traded off with the additional security. The second downside, however, can be resolved by placing another data diode between the audit servers (Fig. 4). In this architecture, the PCN audit server sends its logs to the EN audit server resulting in a unified audit log maintained and accessible from the EN.

4 Net-Centric Cyber Decision Support Systems

The United States Department of Defense (DoD) has adopted the theory of net-centric warfare in its campaign to adapt to warfare in the information age (Gagnon et al. 2010). In this section, we describe an architecture that implements a net-centric cyber decision (NCDS) support system and discuss how data diodes can improve the security and survivability of such systems.

4.1 Information Warfare and OODA Loop

Information warfare is concerned with protecting, improving, and leveraging one's own information while simultaneously corrupting the adversary's information (United States Joint Chiefs of Staff 2006). It comprises two central concepts of *situational awareness* and *decision-making* processes. Decision-making is often conceptualized as a four-stage cycle of observing, orienting, deciding, and acting (*OODA loop*), graphically presented in Fig. 5a. The goal is to achieve decision superiority by making better decisions more quickly.

The cyber OODA loop is the decision-making process enabling cyber-attack defense. It augments cyber survivability in two ways.

1. It facilitates the observation and analysis of the cyber space.

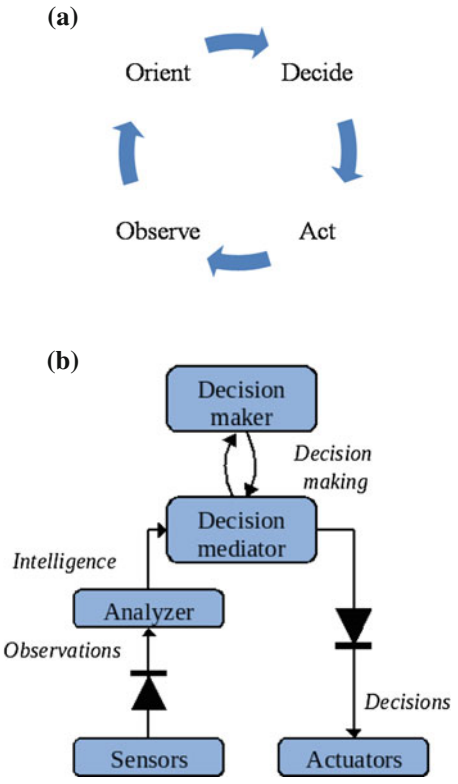

Fig. 5 An NCDS system and its corresponding OODA loop. **a** The OODA loop. **b** An NCDS system with data diodes

2. It enables effective decision-making and fast defensive counter-measure deployment.

4.2 NCDS Architecture

The NCDS system as proposed by Gagnon et al. (2010) utilizes net-centric cyber services to strengthen the cyber OODA loop. In this way, decision makers are empowered to observe the cyber space, make decisions based on the facts, and intelligently respond to cyber attacks. The architecture has six main components (see Fig. 5b):

1. *Cyber sensors* monitor the cyberspace and collect information.
2. *Cyber analyzers* merge sensor data, provide situational awareness, and produce actionable intelligence.
3. *Decision mediators* present situational awareness and potential actions to decision makers in an intelligent manner.
4. *Automated decision-makers* produce "reflex" decisions.

5. *Human decision-makers* produce "cognizant" decisions.
6. *Actuators* enact decisions.

Each of these components can be implemented using distributed services in a service oriented architecture (SOA). For instance, sensors can be implemented on network access devices or on end nodes (desktops and servers). The NCDS architecture works as follows. First, the cyber sensors observe cyber events and collect information. They then pass the information to the *Analyzers* which coalesce sensor data and produce intelligence. Next, the intelligence is passed to the *Mediators* which intelligently select the information to present to the *Decision-makers* and provide a list of potential actions. There can be two types of Decision-makers in the NCDS system: automated Decision-makers which provide quick and safe decisions and human decision-makers who make more complex, potentially dangerous decisions based on more comprehensive contextual facts.

To clarify, consider as an example that the critical infrastructure is under a massive denial-of-service (DoS) attack by servers in a foreign nation. In this case, the sensors which are in the routers and network devices record a sharp increase in traffic volume. The Analyzers, based on the information from various sensors determine that the system is under DoS attack from specific subnets. The Analyzers may also suggest mitigatory actions, for example, a list of filters that can be applied to the routers. The Automated decision-makers, using black listing techniques, apply the filters to the main routers in the network to prevent the traffic from reaching the target system. The filters are then sent to the routers and border firewalls which are the actuators in this case. Human decision-makers can then make longer term decisions ranging from applying permanent filters to specific flows to criminal pursuit/forensics or more drastic maneuvers.

4.3 NCDS with Data Diodes

For the NCDS system to function correctly, a number of security requirements must be satisfied.

1. The integrity of the sensors, analyzers, mediators, automated decision makers, and actuators must be preserved.
2. Since the intelligence is produced by aggregating and analyzing sensor data, it is often classified. Hence, the confidentiality of the intelligence must also be preserved.
3. It is also critically important to preserve the availability of the analyzers, mediators, automated decision makers, and actuators in the presence of cyber attacks.

The confidentiality of sensor data and decisions sent to actuators are arguably less important since an observer can collect this information by passively monitoring the network.

The NCDS system can be augmented with data diodes at two different points in the architecture. Data diodes can be placed (1) between the sensors and cyber analyzers and (2) between the decision mediators and actuators (see Fig. 5b). The first diode preserves the confidentiality of the intelligence (requirement II) whereas the second diode protects the NCDS system against attacks originating from the actuators, thus improving integrity and availability of the decisions (requirements I and III). It is important to note, however, that the diodes in this architecture do not provide confidentiality of the actions, integrity of the sensor data, or cyber attack protection for the analyzers. Since cyber sensors have to inherently collect data, they must be able to send data to the analyzers, exposing them to attacks. To mitigate this problem, additional traffic limitations and format checks must be put in place to strictly limit the traffic from the sensors to the analyzers to validate, sensor-specific data. Finally, integrity of the sensor data must be ensured by sensor hardening or mitigated by distributing a large number of sensors across the network. The data diodes cannot help in this case.

5 Evaluation

5.1 TPCN with Data Diodes

To evaluate the effectiveness of data diodes in augmenting the security of TPCNs, we employed the Common Attack Pattern Enumeration and Classification (CAPEC) database (CAPEC 2008). CAPEC contains attack patterns along with their descriptions, prerequisites, methods, consequences and mitigation strategies. We consider nine attack categories (with 31 attack patterns), which we believe are meaningful in the PCN context and showcase the differences between traditional PCN, TPCN, and TPCN with data diodes. For example, while buffer overflow attacks are effective against software applications, they are not relevant when evaluating network designs.

We qualitatively evaluate each design against the attack patterns and express the feasibility of each attack as high, medium, low, or not feasible. In this context, "high" means that an attack is performed with little effort and cost; "medium" implies that an attack is still possible but requires expert knowledge and is costly; "low" indicates that an attack is highly unlikely or involves enormous effort, time and/or cost; finally, assuming that the data diodes have no back flow, "not feasible" implies that the attack cannot succeed.

Figure 6 illustrates the results. Considering the 31 total attack patterns, a PCN is vulnerable to nineteen (61.3 %) high, nine (29 %) medium, and three (9.7 %) low feasibility attacks. On the other hand, a TPCN is vulnerable to only two (6.5 %) high feasibility attacks along with nine (29 %) medium and twenty (64.5 %) low feasibility attacks. Finally, a TPCN with data diodes is vulnerable to one (3.2 %) high, three (9.7 %) medium , and nineteen (61.3 %) low feasibility attacks. Eight (25.8 %) attacks are not feasible against the TPCN with data diodes. Note that this is a qualitative

comparison of the architectures; the quantitative assessment of network architectures based on security metrics is an open research problem and is beyond the scope of this study.

5.2 NCDS with Data Diodes

To evaluate the protection provided by data diodes is an NCDS system, we use a different strategy. Since the design of NCDS studied in this chapter is high level and we do not want to limit the study to a specific implementation, we cannot directly evaluate the feasibility of the attack patterns. Instead, we use the confidentiality, integrity, and availability impact (CIA impact) of the attacks described in CAPEC to evaluate the NCDS architecture. CAPEC categorizes the CIA impact of the attacks as high, medium, or low. We study 108 attack patterns with high or medium impacts of which 100 are confidentiality, 101 are integrity, and 62 are availability impacts (note that some attacks have more than one impact).

Figure 7 illustrates the results. The first column represents an NCDS system with no protection; the second column depicts the impact of attacks originating from the sensors on the analyzer in an NCDS with data diodes; the last column shows

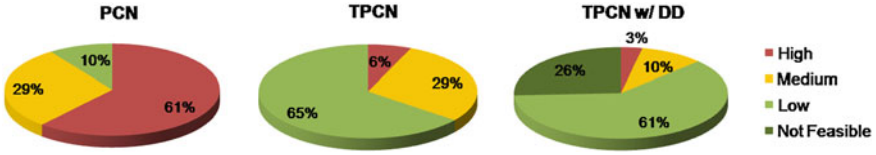

Fig. 6 Feasibility of attack patterns against PCN, TPCN, and TPCN with data diodes

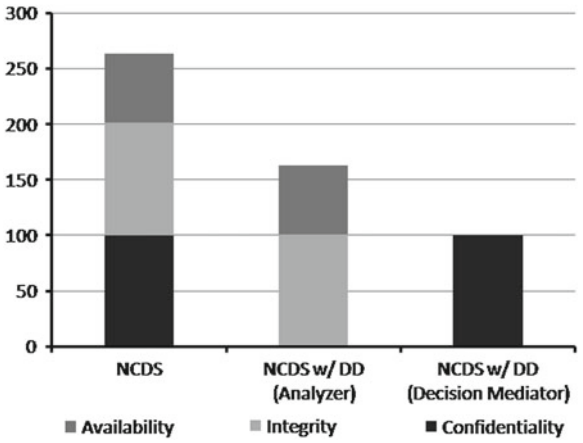

Fig. 7 Attack impacts on NCDS and NCDS with data diodes

the impact of attacks originating from the actuators on the decision mediators. The assessment in Fig. 7 considers the effect of local attacks on individual components of NCDS, assuming that the entire system is not compromised. The diodes provide confidentiality of the intelligence and integrity/availability of the decisions. The integrity/availability of the observations and the confidentiality of the decisions are not provided by data diodes and additional security mechanisms must be deployed to guarantee these properties. However, by simply placing data diodes in an NCDS system, 38 % of the attack impacts on the sensor side and 62 % of the impacts on the actuator side are mitigated. As is the case similar to TPCNs, data diodes can only provide limited guarantees for an NCDS system. Additional mechanisms and infrastructure must be deployed to satisfy all security requirements.

6 Related Work

Kang et al. (2005) first designed and implemented a network device, *network pump*, for limiting covert back flow of data across the network. The network pump keeps the communication bidirectional, but it queues and sends the acknowledgments at probabilistic times. Stevens and Pope (1995) discuss different implementations of data diodes and their assurance levels and limitations. Jones and Bowersox (2006) propose the use of data diodes to implement secure data exports for voting systems. Finally, Roach (2007) demonstrates the application of data diodes in aircraft instrumentation systems. To the best of our knowledge, we are the first to propose the application of data diodes in industrial control and cyber decision support systems, and develop a security infrastructure for their effective deployment and assessment.

7 Conclusion and Future Work

Data diodes can offer some protection at the expense of imposing some limitations to the system. To effectively deploy data diodes within a system the designers must fully understand their functionalities and limitations. Data diodes do not offer a comprehensive security solution, yet they can enhance the security of the system if used with care. TPCN and NCDS are two important critical infrastructure applications that can benefit from the confidentiality or integrity/availability guarantees provided by data diodes. Based on our work on NAD rule conflicts (Okhravi and Nicol 2009), we plan to develop an algorithm to distribute firewall rules in the presence of data diodes in a way that minimizes rule conflicts (Hari et al. 2000) and implement a prototype TPCN and NCDS systems on top of our testbed (Davis et al. 2006).

References

CAPEC (2008) Common attack pattern enumeration and classification

Davis CM, Tate JE, Okhravi H, Grier C, Overbye TJ, Nicol D (2006) SCADA cyber security testbed development. In: Proceedings of the 38th North American power symposium (NAPS 2006), pp 483–488

Gagnon M, Haines J, Kapadia A, Truelove J, Huang O (2010) Towards net-centric cyber survivability for ballistic missile defense. In: 1st international symposium on architecting critical systems federated with CompArch 2010 (ISARCS'10)

Hari A, Suri S, Parulkar G (2000) Detecting and resolving packet filter conflicts. In: Proceedings of IEEE INFOCOM, pp 1203–1212

Hofstadter DR (1979) Godel, Escher, Bach: an eternal golden, 1st edn. Basic Books Inc., New York

Interactive Link Data Diode Device (2010) Manual, BAE Systems

Jones DW, Bowersox TC (2006) Secure data export and auditing using data diodes. In: Proceedings of the USENIX electronic voting technology workshop 2006, EVT'06. USENIX Association, Berkeley, CA, USA, p 4

Kang MH, Moskowitz IS, Chincheck S (2005) The pump: a decade of covert fun. In: Proceedings of the IEEE computer society on 21st annual computer security applications conference ACSAC '05, Washington, DC, USA, pp 352–360

Menoher J, Mraz R (2007) CWID 2007 data diode case study. In: Invited presentation at the 23st annual computer security applications conference (ACSAC '07)

Network Admission Control (NAC) (2005) Technical overview, Cisco Systems, Inc.

Okhravi H, Nicol D (2009) Application of trusted network technology to industrial control networks. Elsevier Int J Crit Infrastruct Prot (IJCIP) 2(3):84–94

Okhravi H, Nicol D (2008) Applying trusted network technology to process control systems. In: Goetz E, Shenoi S (eds) Critical infrastructure protection II, 2nd edn. Springer, Boston, pp 57–70

Rieback MR, Crispo B, Tanenbaum AS (2006) Is your cat infected with a computer virus? In: Proceedings of the fourth annual IEEE international conference on pervasive computing and communications, pp 169–179

Roach J (2007) The architecture of aircraft instrumentation networks. In: Proceedings of the international telemetering conference (ITC 2007)

Stevens M, Pope M (1995) Data diodes. Technical report DSTO-TR-0209, Electronics and Surveillance Research Laboratory (DSTO)

United States Joint Chiefs of Staff (2006) Joint publication, information operations, pp 3–13. http://www.dtic.mil/doctrine

Waterfall's Unidirectional Security Gateways (2010) Manual, waterfall. http://www.waterfallsecurity.com/technology/

Wool A (2004) A quantitative study of firewall configuration errors. Computer 37(6):62–67

Index

A
Actuators, 212
Agent, 104, 106, 109, 114
AMI, 164, 166, 172
Ancillary services, 84, 85, 94
Appropriate use banners, 150
Architecture, 163, 165
Attack vectors, 107, 110
Audit log performance impact, 146, 151
Audit record contents, 144, 148
Audit record retention, 148
Authentication, Authorization, Access Control (AAA) Server, 207
Automated decision-makers, 211
Automated meter reading (AMR), 111, 166
Availability, 205

B
Benefits of DR, 82, 85
Black-box instrumentation strategy, 110
Boundary protection, 145, 150
BPL, 166
Branch-and-bound, 13, 14, 23

C
Capacity markets, 84, 85
Centralized approach, 9, 21
Central pattern generator (CPG), 44, 46
Challenges, 174, 175
Circuit stability, 45, 46
Comisión federal de electricidad (CFE), 126, 128
Common Attack Pattern Enumeration and Classification (CAPEC) database, 213
Common object request broker architecture (CORBA), 122
Communication confidentiality, 158
Communication integrity, 146, 157, 158
Communications, 162, 166, 171, 175
Confidentiality, 205
Confidentiality, integrity and availability impact (CIA impact), 214
Confidentiality of information at rest, 158
Cost-effectiveness, 56, 67, 69
Cost of implementation, 82, 83
Critical peak pricing, 83, 84, 86, 91–93
Customer, 55–57, 59, 77
Cyber analyzers, 211
cyber decision support systems, 215
Cyber security, 168–170
Cyber sensors, 211

D
Danger-associated molecular patterns (DAMPs), 44
Decision-making, 210
Decision mediators, 211
Decomposition, 3, 10, 14, 15, 23, 26, 28, 29
Define auditable events, 144, 148
Degeneracy, 45, 50
Demand bidding programs, 84, 85
Demand response (DR), 80–84, 86, 91, 93
Demand response objectives, 79, 80, 81, 87, 94
Demand response potential, 79, 80, 82, 86–88, 91–93
Demand shifting, 81, 92
Demand side management (DSM), 80, 81, 93, 94

V. Pappu et al. (eds.), *Optimization and Security Challenges in Smart Power Grids*, Energy Systems, DOI: 10.1007/978-3-642-38134-8,

Denial of service attack (DOS attack), 135, 137, 212
Denial-of-service protection, 144, 151
DER, 164, 166–168
Device identification and authentication, 157
Diagnostic support, 48, 49
Differentiated service code point (DSCP), 129
Direct load control, 83, 85
Directory server, 208
Distributed algorithm, 22
Distributed generation, 56–59
Distribution grid, 53, 58, 61
Double sectionalization, 63, 68, 75, 76
DR mechanism, 80–83, 86, 88, 94, 95
Duality, 13, 23

E
Eavesdropping, 134
Economic dispatch, 2–5, 7, 9, 11, 13, 15, 17, 19–21, 23–25, 27, 29–31, 33, 35, 37
Emergency demand programs, 83, 85
Enforce access authorizations, 145, 149
Enterprise network (EN), 206
Ethernet, 121, 126, 128
Event sequences, 103, 106, 107, 114
Exact algorithm, 10, 11

F
Fault, 55, 56, 62, 64
Fault clearing, 53
Fault clearing cost, 64
Feeder, 54–56, 58, 60, 62, 64
Feeder reconfiguration, 55, 57, 58, 60, 61
Feeder section, 54
Feeder sectionalization, 55, 61
Flexible load shape, 81

G
Generic object orientated substation event (GOOSE), 130
Global optimization, 21, 23
Global positioning system (GPS), 127

H
Heuristic algorithm, 3, 10, 11, 18
Home appliances, 87, 88, 91
Homeostatic control, 39, 47
Human decision-makers, 212
Hypothalamic-pituitary-adrenal (HPA) axis, 42

I
IEC 61850, 117, 119–121, 123, 124, 126, 128, 132, 134, 137
IEEE 1588, 127
Incentive-driven DR mechanism, 83–85
Independent supply point, 56, 61, 62
Industrial control, 215
Industrial control systems, 205
Inflammasomes, 44, 47
Innate immunity, 41–43
Intelligent electronic device (IED), 121–124, 126, 127, 131, 133–135
Interior-point method, 21, 22, 27
Internet engineering task force (IETF), 130
Interoperability, 162, 168, 169
Interruptible programs, 83, 86, 93
Intrusion detection, 151
Intrusion detection system (IDS), 125
Investment-to-sectionalization efficiency, 53, 56, 73, 76
Islanding, 47, 49

L
Lagrangian, 3, 10, 12, 13, 15, 18, 19, 22, 23, 26
Least privilege, 145, 149
Load, 59, 60, 72, 74
Load shedding, 47, 49

M
Macrophages, 43, 44
Master terminal unit, 206
Message authenticity, 146, 157
Method-applicable grid, 62, 65, 69
Microgrids, 39, 40, 49
Mixed integer programming, 14, 19
Monitoring, 151, 154
Multi-agent systems (MAS), 47, 49
Multilink point-to-point protocol (MLPPP), 130

N
NASPI, 167
Net-centric cyber decision (NCDS) support system, 204, 210

Network access device, 207
Network pump, 215
Network time protocol, 127
Neural control, 39, 42, 47, 50
Non-distributed energy, 54, 56, 62, 64
Non-distributed energy cost, 53, 62, 66, 68, 72, 75, 76
Non-repudiation, 146, 157
North America Electric Reliability Corporation (NERC), 105

O
OODA loop, 210
Open systems interconnection (OSI), 121
Operational security, 39–41, 45, 49
Operator's interruption cost, 53, 64, 77
Optimality, 12, 15, 22
Out-of bounds communication, 206

P
Parallel, 3, 18, 27, 29
Password complexity, 145, 149
Patch management, 145, 149
Pathogen-associated molecular patterns (PAMPs), 43, 44
Peak shaving, 81, 82, 87, 89
Phasor, 167
PHEV, 166−168
PMU, 164, 167
Ports and services, 142, 144, 145, 148, 150
Post-contingency, 25, 26
Posture remediation servers, 208
Posture validation servers, 207
Power flow, 2, 4, 5, 8, 14, 26, 28
Power system restoration, 48, 49
Previous logon notification, 145
Price-driven DR mechanism, 83, 84
Primary control, 48
Priority list, 11, 12
Process bus, 123
Process control network, 204, 206
Programmable logic controllers, 206
Program instrumentation, 109, 110
Protection of audit information, 157

Q
Quality of service (QOS), 129

R
Real time pricing, 83, 85, 86
Reason-based sectionalization, 53, 75, 76
Recloser, 54–56, 62, 67
Recloser placement, 53, 60, 61, 76
Recovery and reconstitution, 146, 151
Redundancy, 45
Remote terminal units, 206
Renewable, 1–3, 30
Replay attack, 134
Robust, 16, 19, 29, 30

S
SAIDI, 59, 60, 65, 76
SAIFI, 59, 60, 76
SCADA, 100–102, 105, 107, 110, 114, 134, 164, 169, 173
SCADA systems, 48, 49
Secondary control, 48, 49
Sectionalizing point, 62, 65, 68, 72, 76
Security, 162, 168, 170, 171, 174, 175
Semidefinite programming, 23
Sensors, 167
Service oriented architecture, 212
Session lock, 150
Single sectionalization, 63, 65, 68, 75, 76
Smart grids, 54, 61, 76, 169, 172
Smart grids security, 165, 168–170, 175
Smart meter, 100
Spinning reserve, 2, 15, 16
Standards, 162, 168, 169, 175
Station bus, 123, 124
Substation, 39, 46, 47, 49, 164, 165, 170, 172, 173
Substation automation system (SAS), 120
Substation configuration language (SCL), 120, 125
Substations, 46–49
Supervisory control and data acquisition (SCADA), 124
Support vector machine (SVM), 125
Survey, 175

T
Tennessee valley authority (TVA), 125
Threat models, 110
Time of use pricing, 85
Time stamp, 144, 148

TPCN architecture, 207
Triangulation, 114
Trusted path, 146, 157
Trust-region, 21, 22
Two-stage, 17, 19, 28

U
Uncertainty, 1, 3, 4, 15, 17–19, 24, 25, 30
Unidirectional communication, 204
Unit commitment, 2–5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37
Unsuccessful login attempts, 145, 150
Use of validated cryptography, 158

V
Vagus nerve, 42
Valley filling, 81, 92
Voting systems, 215
Vulnerability assessment, 143, 145

W
White-box instrumentation strategy, 110