Security Games and Risk Minimization for Automatic Generation Control in Smart Grid

Yee Wei Law, Tansu Alpcan, Marimuthu Palaniswami, and Subhrakanti Dey

Dept. of Electrical & Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia
{ywlaw,tansu.alpcan,palani,sdey}@unimelb.edu.au

Abstract. The power grid, on which most economic activities rely, is a critical infrastructure that must be protected against potential threats. Advanced monitoring technologies at the center of smart grid evolution increase its efficiency but also make it more susceptible to malicious attacks such as false data injection. This paper develops a game-theoretic approach to smart grid security by combining quantitative risk management with decision making on protective measures. Specifically, the consequences of data injection attacks are quantified using a risk assessment process based on simulations. Then, the quantified risks are used as an input to a stochastic game model, where the decisions on defensive measures are made taking into account resource constraints. Security games provide the framework for choosing the best response strategies against attackers in order to minimize potential risks. The theoretical results obtained are demonstrated using numerical examples.

Keywords: Smart grid, automatic generation control, security games

1 Introduction

A power grid is a critical infrastucture that must be protected against potential threats. As it evolves to a "smart grid" with better efficiency, however, the security concerns increase due to emergence of new attack vectors exploiting evolving system complexity. While security is an important issue for grid operators, real world constraints such as resource limitations necessarily force adoption of a risk management approach to the problem. Protective measures are usually taken based on a cost-benefit analysis balancing available defensive resources with perceived security risks. This paper investigates the important class of false data injection attacks to smart grids which directly affect the operation of automatic generation control systems and potentially lead to blackouts. The problem is formulated first as one of quantitative risk management which in turn is used as an input to a stochastic (Markov) security game. The resulting game analysis helps smart grid operators to make informed decisions on their security strategies while taking into account their resource constraints. Although the paper focuses on a certain type of attack and subsystem, the approach can be applied to similar security problems in smart grid, and hence, can be extended to develop the foundation of a systematic framework for smart grid security.

A simple but elegant definition of risk is "the probability and magnitude of a loss, disaster, or other undesirable event" [14]. **Security risk analysis** can be defined as "the process of identifying the risks to system security and determining the likelihood of occurrence, the resulting impact, and the additional safeguards that mitigate this impact" [27]. Most smart grid standards and guidelines (e.g., IEC 62351-1, NISTIR 7628) identify risk assessment as a critical part of a security framework. For instance, the Australian Government advocates the use of the Australian and New Zealand Standard for Risk Management (AS/NZS ISO 31000:2009) by owners and operators of critical infrastructure [4]. However, the standard ISO 31000:2009 is "not mathematically based", and has "little to say about probability, data, and models" [19]. Comprehensive risk assessment is hampered by the following trends:

Stealthy attacks A zero-day vulnerability is a vulnerability exploited by some malware before or on the same day it is known by the vendor. Stuxnet–the world's first computer worm that targets programmable logic controllers–exploited as many as four zero-day vulnerabilities, allowing it to spread undetected by commercial antimalware software. In a 2011 report, McAfee found the electric sector has the highest occurrence of Stuxnet among the power, oil, gas and water sectors [5]. Stuxnet is a classic example of a malware developed with nation-state resources. The discovery of Stuxnet successors Duqu and Flame suggests comprehensive risk assessment must go beyond detectable attacks that target ICT systems to stealthy attacks that target control systems.

Forever-day vulnerabilities Power control systems were not originally designed to be connected to the Internet, and thus lack many of the security controls found on corporate IT systems. Some experts estimate current control system security to be a decade behind enterprise IT security [28]. As more power control systems become connected to corporate networks, it is increasingly possible for Internet-originated attacks to penetrate power control systems through corporate networks. The bad news is that control system vendors are refusing to patch legacy systems, giving rise to "forever-day vulnerabilities" [12].

Complexity Power grids are complex systems, and the global drive toward *smart grids* is making existing systems even more complex. The 47 actors and 137 inter-actor interfaces identified in NIST's logical reference model of a smart grid [26] present a large attack surface with no shortage of entry points. Risk assessment methodologies that rely on expert judgements, when no one expert can claim full unbiased knowledge of even a small part of the system, are error-prone. To assess and mitigate the security risks faced by power control systems, a systematic approach that is based on empirical evidence is clearly needed.

Security games provide an analytical framework for modeling the interaction between malicious attackers, who aim to compromise smart grid, and operators defending them. The "game" is played on smart grids, which are complex and interconnected systems. The rich mathematical basis provided by the field of game theory facilitates formalising the strategic struggle between attackers and defenders for the control of the smart grid [1]. Utilising the risk framework and some of the concepts of earlier studies [7, 24], this work applies game theory to

the modeling of attacks on and defenses for a critical power system component called automatic generation control.

Fig. 1. An overview of the methodology adopted in this paper.

The **main contributions** of this work include

- Assessment and identification of risks faced by the automatic generation control system, which constitute an important part of smart grid.
- A discussion of the security threat model, potential attacks, and countermeasures.
- A quantitative risk model capturing the probability and magnitude of security threats faced by the automatic generation control system due to false data injection attacks.
- A stochastic (Markov) security game for analysis of best defensive actions building upon the risk analysis conducted and under resource limitations.
- A numerical study illustrating the framework developed.

The rest of the paper is organized as follows. Section 2 states the problem of assessing the cyber security risks of automatic generation control, an essential power system component. Section 3 defines the threat model. Section 4 discusses attack and defense actions within this threat model. Section 5 presents our game and risk model. In Section 6, we apply the game and risk model to automatic generation control, and present our simulation results. Section 7 discusses related work, and finally Section 8 concludes this paper.

2 Problem statement: automatic generation control (AGC)

The most critical aspect of a power system is stability, and one of the most important parameters to stabilize is frequency. This is because the frequency of a power system rises/falls with decreased/increased loading. Failure to stabilize frequency may lead to damage to equipment (utility's or end users'), harm to human safety, reduction of or interruption to electricity supply. Violation of frequency stability criteria is one of the main reasons for numerous power blackouts [6]. Less tangible secondary impacts, including loss of data or information and damage to reputation, are equally undesirable.

The frequency control system operates at three levels. Primary frequency control takes the form of a turbine governor's *speed regulator*, a proportional controller of gain $1/R$, where R is the *droop characteristic* (drop in speed or frequency when combined machines of an area change from no load to full load). Secondary frequency control is for correcting the steady-state error residue

left by the proportional controller, and may take the form of an integral controller; in which case, primary and secondary frequency control form a parallel proportional-integral controller, capable of driving frequency deviations to zero whenever a step-load perturbation is applied to the system. Tertiary frequency control is supervisory control based on offline optimizations for (i) ensuring adequate spinning reserve in the units participating in primary control, (ii) optimal dispatch of units participating in secondary control, (iii) restoration of bandwidth of secondary control in a given cycle. While primary and secondary control respond in seconds and tens of seconds respectively, tertiary control is usually manually activated minutes after secondary control. Our study concerns only the *dynamics* of frequency control, and hence does not consider tertiary control.

In an interconnected system with two or more *control areas*, in addition to frequency, the generation within each area must also be controlled to maintain scheduled power interchanges over *tie lines* (inter-area transmission lines). The control of both frequency and generation is called *load-frequency control*. Within each area, each generation unit has primary control, while secondary control is centralized. Together, decentralized primary control and centralized secondary control achieve the purpose of load-frequency control. *Automatic generation control* (AGC) is load-frequency control with the additional objective of *economic dispatch* (distributing the required change in generation among units to minimize costs) [18, 38]. However, AGC is sometimes referred to as automated (vs manual) load-frequency control [3], or even the entire frequency control system itself [23]. AGC is an indispensable part of the "central nervous system" of a power grid called the *energy management system* (EMS), and possibly the only automatic closed loop between the IT and power system of a control area [10]; because of this, it is subject to attacks propagated through the IT system. A detailed threat model is given in Section 3.

When system frequency deviates from the nominal frequency (60 Hz for Americas, 50 Hz for most other parts of the world) by a certain threshold, overfrequency and underfrequency protection relays execute tripping logic defined by a protection plan that varies from operator to operator. Assuming a nominal frequency of 60 Hz, overfrequency relays start tripping thermal plants when frequency rise exceeds 1.5 Hz [22, 23], but these relays are usually set to tolerate deviations due to post-fault transients for short periods of time. Underfrequency relays perform *underfrequency load shedding* (UFLS), which is the sole concern of our study because it results in directly measurable revenue loss. For our study, we adopt Mullen's UFLS scheme [25]. In Algorithm 1, Δf denotes frequency deviation. $\Delta P_{\text{safe}} \stackrel{\text{def}}{=} -0.3/R$, where R is the droop characteristic of the generators. $\Delta P_{\text{est}} \stackrel{\text{def}}{=} \Delta P_m - \Delta P_e$, i.e., change in mechanical power minus change in electrical power.

Algorithm 1: Mullen's UFLS scheme [25]
Sampling time: 0.05 s. Nominal frequency: 60 Hz. *Not* for overfrequency protection.

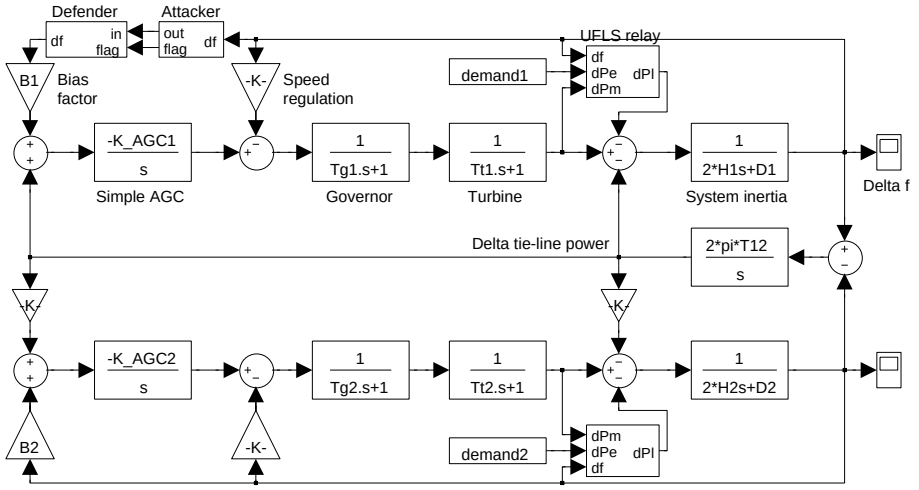
if timer == 0 **then**
  if $\Delta f \leq -0.4$ **and** $\Delta P_{\text{est}} + \Delta P_{\text{safe}} > 0$ **then**
    timer $\leftarrow$ 1 // Level-1 alarm

$P_{\text{sched}} \leftarrow \Delta P_{\text{est}} + \Delta P_{\text{safe}}$
else if $-0.4 < \Delta f \leq -0.35$ **and** $\Delta P_{\text{est}} + \Delta P_{\text{safe}} > 0$ **then**
timer $\leftarrow$ 2 // Level-2 alarm
$P_{\text{sched}} \leftarrow \Delta P_{\text{est}} + \Delta P_{\text{safe}}$
else if UFLS is in effect **and** $-0.35 < \Delta f \leq 0$ for some time **then**
Reconnect most recently shed loads
end
return
end
if timer > 0 **then**
timer $\leftarrow$ timer - 1
if timer $== 0$ **then**
Shed P_{sched}
end
end

Our study is based on the two-area AGC system model and associated simulation parameters in Fig. 2, which incorporates a simple turbine-governor model [6]. The automatic generation controller is an integral controller of gain K_{AGC}. We note that design of AGC is an established area with designs dating back to the 1950s; [15] alone surveys over a hundred designs. A simple integral controller seems to be a logical starting point. Following convention, we model the AGC system as continuous-time. We set the nominal frequency to 60 Hz. The demand time series `demand1` and `demand2` are the demand profiles of Victoria on 4-5 June 2012 and of South Australia on 7-8 June 2012 respectively, provided by the Australian Energy Market Operator. The UFLS relays in both areas execute Algorithm 1 every 0.05 s. Once the system frequency has stabilized for at least 30 s, the UFLS relays reconnect the shed loads in the reverse order they were shed. The maximum sheddable loads are capped at 4 p.u. and 1 p.u. for areas 1 and 2 respectively. "p.u." stands for "per unit" and is simply the ratio of an absolute value in some unit to a base/reference value in the same unit. The base load for both areas is taken to be 1000 MW (hence 4 p.u. is 4000 MW in this case).

3 Threat model

Access to a control system is typically enabled through a *virtual private network* (VPN) [37]. As VPN is usually the only access control mechanism [33], gaining unauthorized access to a control system is no different from infiltrating any IT network. Threats to control systems are well documented [32]. VPNs offer no resistance to insider attackers who possess the required access rights, either in the form of passwords or physical access to SCADA network terminals. An often-used attack vector by outsider attackers is a Trojan a system operator unknowingly downloads when he/she visits a malicious web site or opens an infected email attachment. By logging keystrokes, stealing private keys, etc., the Trojan captures the necessary access credentials for the attacker. Based on



Symbols	Definition	Symbols	Definition
H, D	Inertia constant, damping coefficient	T_g, T_t	Time constants for turbine governor
R	Droop characteristic	B	Frequency bias factor
K_{AGC}	Gain of secondary frequency control	T_{12}	Tie-line synchronizing coefficient

	$K_{\mathrm{AGC}\,i}$ (s)	D_i (p.u./Hz)	$2H_i$ (p.u. s)	R_i (Hz/p.u.)	T_{gi} (s)	T_{ti} (s)	B_i (p.u./Hz)
$i = 1$	0.3	0.015	0.1667	3.00	0.08	0.40	0.3483
$i = 2$	0.2	0.016	0.2017	2.73	0.06	0.44	0.3827

Fig. 2. Simulink representation and simulation parameters for a two-area AGC system model based on Bevrani's [6, Fig. 2.10 and Table 2.2]. The top area is labeled area 1.

information from multiple sources [9, 33, 37, 38], Fig. 3 shows the typical communication architecture of a control center and a substation. Some authors [34] equate the compromise of an entire control center or substation to the successful cracking of a VPN access password and the penetration of an Internet-facing firewall in Fig. 3; this strong attacker model is not entirely unrealistic, but our goal is to investigate the strategy of an attacker that has successfully penetrated the VPN but whose actions within the AGC system are bounded by several resource constraints. We assume the following resource constraints:

- The attacker cannot directly trip generators, or transmission lines (by opening circuit breakers).
- The attacker cannot tamper with turbine governors.
- The attacker cannot tamper with underfrequency load shedding (UFLS) relays. Some commercial relays (e.g., SEL-387E) have an integrated frequency meter, and are thereby not subject to false frequency data injection attacks.
- The attacker cannot tamper with the EMS.
- The attacker can reduce but not block the input/output of the EMS.

Without the above constraints, it is a trivial exercise for any attacker that has successfully penetrated the VPN to trigger cascading failures across the power grid. It is therefore conceivable that an energy provider would make protecting

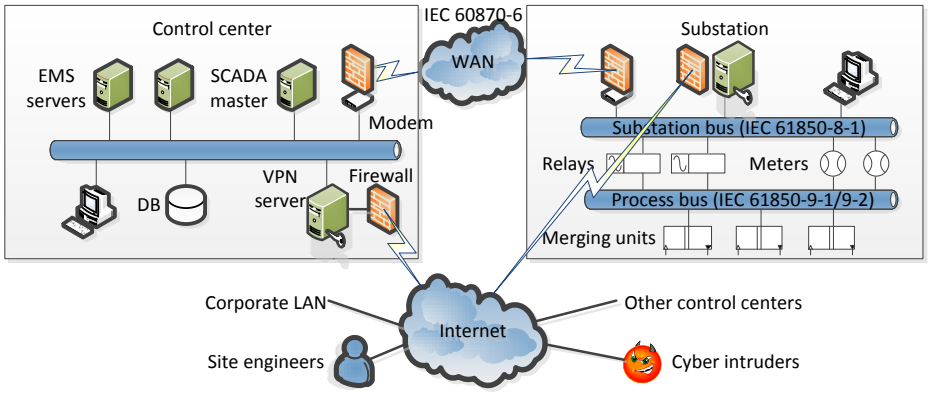


Fig. 3. Accessibility of a power system control center and substation from the Internet. AGC is executed on one of the EMS servers. In our threat model, an attacker can feed the AGC software with false frequency data.

its generators, circuit breakers, turbine governors, UFLS relays, and EMS its foremost priority. Despite the above constraints, an attacker can forge and send false frequency data to the AGC software executing on one of the EMS servers, by masquerading as one of the relays (except the UFLS relays) or meters in the substation (see Fig. 3). In the spirit of stealthy attacks as embodied by Stuxnet, Duqu and Flame, it is also conceivable that a persistent attacker would adopt this subtle and stealthy strategy. It is up to the AGC software to detect this attack. False data attacks on the speed regulator (primary frequency control) are not considered because the machine is usually directly wired to a frequency sensor without going through a communication network. In the next section, several potential injection attacks, defense actions and their effects are discussed.

4 Attacks and defense actions

It is impossible to simulate all data injection attack scenarios, but there are three basic attack types on which more sophisticated attacks can be based.

Constant injection If an attacker injects a constant false value, then the it effectively disables the integral control loop, causing the system frequency to converge to a non-nominal frequency. If the false value is positive, then the system will settle on a below-nominal frequency, causing loads to be shed; otherwise, the system will settle on an above-nominal frequency, causing generators to be tripped. Both cases lead to cascading failures.

Overcompensation If an attacker injects a false frequency k times the true frequency, where k is a large positive number, then it effectively causes overcompensation by the integral control loop, and consequently unstable oscillations. As the system frequency sweeps past the overfrequency and underfrequency thresh-

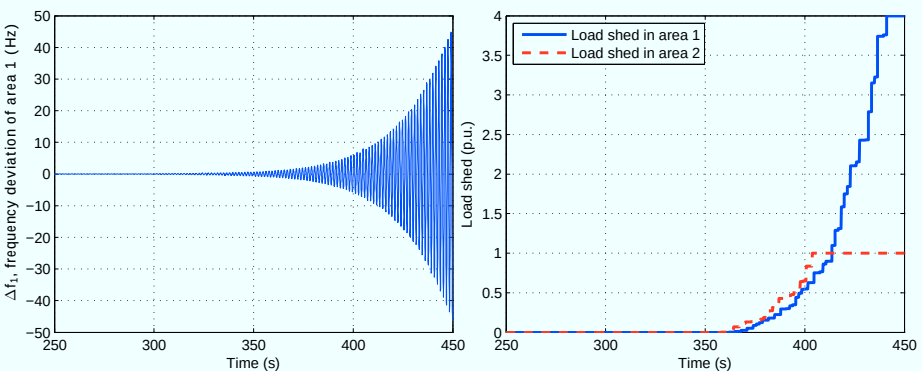


Fig. 4. An example of "overcompensation" attack, where the attacker substitutes Δf_1 with $8\Delta f_1$ as frequency input to the area-1 integral controller. As long as the attack persists, neither generator tripping nor load shedding helps stabilize the system.

olds, generators will be tripped and loads will be shed, followed by cascading failures. Fig. 4 shows the result of an attack using $k = 8$.

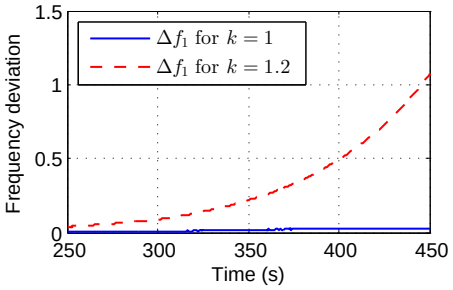


Fig. 5. Negative compensation attack: for large enough k (e.g., 1.2), the system frequency $\to +\infty$.

Negative compensation If an attacker injects a false frequency $-k$ times the true frequency, where k is a positive number, then it effectively reverses the intended effect of the integral control loop, causing the system frequency to diverge from the nominal frequency (see Fig. 5). This attack directly triggers generator tripping, but not load shedding.

For our study, we concentrate only on the overcompensation attack, as it inflicts maximum damage in terms of triggering both load shedding and generator tripping (although we do not simulate the latter). It is also harder to detect than constant injection. Here, we outline some defenses against the overcompensation attack. The first observation is that we can constrain the attack by limiting the frequency input to the integral controller to $[-4.5, 3.5]$ Hz (i.e., passing the input through a saturation filter), because at $\Delta f = -4.5$ Hz, not only should all sheddable loads have been shed, but also all generators would be tripped; at $\Delta f = 3.5$ Hz, all generators would be tripped as well [22]. A common security

measure is redundancy. Multiple frequency meters of different builds can be installed, so that the likelihood of all meters being compromised is small and the AGC software has a non-zero chance of receiving genuine frequency data.

There are unlimited ways to improve upon the overcompensation attack to counter the above defenses. Correspondingly, there are unlimited ways to detect these improved attacks with varying accuracy, and certainly there are more advanced controllers that are less susceptible to these attacks. Nevertheless, our interest is not on the design of attacks, defenses or controller, but on the modeling of system risk dynamics under the actions of the attacker and defender.

5 Game and risk model

Our model is based on Alpcan and Başar's framework [1]. The concept of *risk states* is central to this model. A system has a set of states, and a different level of risk is associated with each state. In this work, we define risk as the product of the probability of a successful attack and the resultant shed load. Clearly under this definition, risk ranges from 0 to the maximum sheddable load for all areas combined, but we partition this risk space into only two states: s_0 where risk is zero (no load is shed), and s_1 where risk is nonzero (some load is shed). We model the state to evolve probabilistically according to a defined stochastic process with the Markov property. Accordingly, we model the interactions between an attacker and a defender using stochastic or Markov *security games*.

As a general basis for Markov security games, consider a 2-player (attacker vs. defender) zero-sum Markov game played on a finite state space, where each player has a finite number of actions to choose from. Let the attacker's action space be $\mathcal{A}^A \stackrel{\text{def}}{=} \{a_1, \ldots, a_{N_A}\}$, the defender's action space be $\mathcal{A}^D \stackrel{\text{def}}{=} \{d_1, \ldots, d_{N_D}\}$, and the state space be $\mathcal{S} \stackrel{\text{def}}{=} \{s_1, \ldots, s_{N_S}\}$. It is assumed that the state evolves according to a discrete-time finite-state Markov chain which enables utilization of well-established analytical tools to study the problem. Then, the *state transitions* are determined by the mapping $\mathcal{M} : \mathcal{S} \times \mathcal{A}^A \times \mathcal{A}^D \to \mathcal{S}$. Let $\boldsymbol{p}^{\mathcal{S}}(t)$ be the probability distribution on the state space $\mathcal{S}$, i.e.,

$$\boldsymbol{p}^{\mathcal{S}}(t) \stackrel{\text{def}}{=} \left[\Pr[s(t) = s_1] \Pr[s(t) = s_2] \cdots \Pr[s(t) = s_{N_S}]\right]^T,$$

where $t \geq 1$ denotes the discrete time (stage) of the repeated Markov game. The mapping $\mathcal{M}$ can then be represented by the $N_S \times N_S$ state transition matrix $\boldsymbol{M}(a, d) = [M_{s_i,s_j}(a, d)]_{N_S \times N_S}$, which is parameterized by $a \in \mathcal{A}^A$ and $d \in \mathcal{A}^D$, such that

$$\boldsymbol{p}^S(t+1) = \boldsymbol{M}(a, d)\boldsymbol{p}^S(t). \tag{1}$$

The matrix entry $M_{s_i,s_j}(a, d)$ represents the probability of state s_i transitioning to state s_j under attacker action a and defender action d.

The mapping $\mathcal{M}$ can alternatively be parameterized by the state to obtain as many zero-sum game matrices $\boldsymbol{G}(s)$ as the number of states, each of dimension $N_A \times N_D$. In other words, given a state $s(t) \in \mathcal{S}$ at a stage t, the players

play the zero-sum game $\boldsymbol{G}(s(t)) = [G_{a,d}(s(t))]_{N_A \times N_D}$. The matrix entry $G_{a,d}(s)$ represents the attacker's gain from risk state s by taking action a when the defender action is d. Using our definition of risk in this work, $G_{a,d}(s)$ is the total load shed in state s under attacker action a and defender action d. By definition, $\boldsymbol{G}(s_0) = \mathbf{0}$. In zero-sum Markov games, the attacker's gain (loss) equals the defender's loss (gain).

The attacker's strategy is defined as a probability distribution on $\mathcal{A}^A$ for a give state s, i.e., $p^A(s) \stackrel{\text{def}}{=} \left[\Pr[a(s) = a_1] \cdots \Pr[a(s) = a_{N_A}]\right]^T$. The defender's strategy is similarly defined. For the zero-sum Markov game formulation here, the defender aims to minimize own aggregate cost, $\bar{Q}$, in response to the attacker who tries to maximize it. The reverse is true for the attacker due to the zero-sum nature of the game. Hence, it is sufficient to describe the solution algorithm for only one player, which is the defender in our case.

The game is played in stages over an infinite time horizon. As in Markov Decision Process, the aggregate cost of the defender at the end of a game is the sum of all realized stage costs discounted by a scalar factor $\alpha \in [0, 1)$:

$$\bar{Q} \stackrel{\text{def}}{=} \sum_{t=1}^{\infty} \alpha^t G_{a(t),d(t)}(s(t)), \qquad a(t) \in \mathcal{A}^A, d(t) \in \mathcal{A}^D, s(t) \in \mathcal{S}, \tag{2}$$

where $G_{a(t),d(t)}(s(t))$ is the $(a(t), d(t))$-th element of the stage-t game matrix $\boldsymbol{G}(s(t))$. The defender can theoretically choose a different strategy $p^D(s(t))$ at each stage t of the game to minimize the final realized cost $\bar{Q}$ in (2). Fortunately, this complex problem can be simplified significantly. First, it can be shown that a stationary strategy $p^D(s) = p^D(s(t)), \forall t$ is optimal, and hence there is no need to compute a separate optimal strategy for each stage. Second, the problem can be solved recursively using *dynamic programming* to obtain the stationary optimal strategy (solving a zero-sum matrix game at each stage). Unlike Markov Decision Process, the optimal strategy can be mixed, i.e., stochastic for each state s. At a given stage t, the optimal cost $Q_t(a, d, s)$ (called Q values) can be computed iteratively using the dynamic programming recursion

$$Q_{t+1}(a, d, s) = G_{a,d}(s) + \alpha \sum_{s' \in \mathcal{S}} M_{s,s'}(a, d) \cdot \min_{p^D(s')} \max_{a} \sum_{d \in \mathcal{A}^D} Q_t(a, d, s') p_d^D(s'), \tag{3}$$

for $t = 0, 1, \ldots$ and a given initial condition Q_0. In (3), $p_d^D(s')$ is the element of $p^D(s')$ that corresponds to d. (3) converges to the optimal Q^* as $t \to \infty$.

There are multiple ways to implement (3). The algorithm called *value iteration* is prescribed here due to its scalability. To describe the algorithm, we first split (3) into two parts:

$$V(s) = \min_{p^D(s)} \max_{a} \sum_{d \in \mathcal{A}^D} Q_t(a, d, s) p_d^D(s), \tag{4}$$

$$Q_{t+1}(a, d, s) = G_{a,d}(s) + \alpha \sum_{s' \in \mathcal{S}} M_{s,s'}(a, d) V(s'), \quad t = 1, 2, \ldots \tag{5}$$

(4) can further be formulated as a linear program:

$$\begin{aligned} &\min_{p^D(s)} V(s) \\ &\text{s.t.} V(s) \leq \sum_{d \in \mathcal{A}^D} Q_t(a,d,s) p_d^D(s), \forall a \in \mathcal{A}^A, \\ &p_d^D \geq 0, \sum_d p_d^D = 1, \forall d \in \mathcal{A}^D. \end{aligned} \tag{6}$$

The strategy $p^D(s), \forall s \in \mathcal{S}$ computed from (6) is the *minimax* strategy w.r.t. Q. The fixed points of equations (4) and (5), V^* and Q^*, lead to the optimal minimax solution for the defender. The value iteration algorithm, using (4), (5) and (6) to find V^* and Q^*, is given in Algorithm 2.

Algorithm 2: The value iteration algorithm

Given arbitrary $Q_0(a,d,s)$ and $V(s)$
repeat
  for $a \in \mathcal{A}^A$ and $d \in \mathcal{A}^D$ **do**
    Update V and Q according to (4) and (5)
  end for
until $V(s) \to V^*$, i.e., $V(s)$ converges

6 Security game and simulation results

A reasonable definition of risk is the product of the probability of a successful attack and the resultant shed load. We define two risk states, i.e., we partition the relative risk probability simplex into two risk regions: s_0 where no load is shed, and s_1 where some load is shed. In the absence of attacks or large disturbances, the system only operates in state s_0.

In our security game, the AGC software reads N consecutive samples alternately from two frequency meters of different builds (one is more secure than the other). N consecutive samples from one meter constitute one *session/stage* (see Fig. 6(a)). The attacker can perform the following actions:

a_1 Send N samples, $N/2$ of which are false.
a_2 Send N samples, N of which are false.

a_1 and a_2 are two special cases of the general attack action space $\mathcal{A}^A = \{$ Send N samples, i of which are false $(i = 1, \ldots, N)\}$. We consider 2 out of N possible attack actions merely for numerical simplicity. The attacker sets a *false sample* to -4.5 Hz if the true Δf is negative, or 3.5 Hz if the true Δf is positive. This implements the overcompensation attack, and takes into account the saturation filter in Section 4. Correspondingly, the defender can perform the following actions:

d_1 Upon collecting N samples, run Detection Algorithm 1 (to be defined later). If detection result is positive, disinfect the meter (e.g., by refreshing its firmware, cryptographic keys and so on). Disinfection is assumed to complete within the time of one session (see Fig. 6(a)).

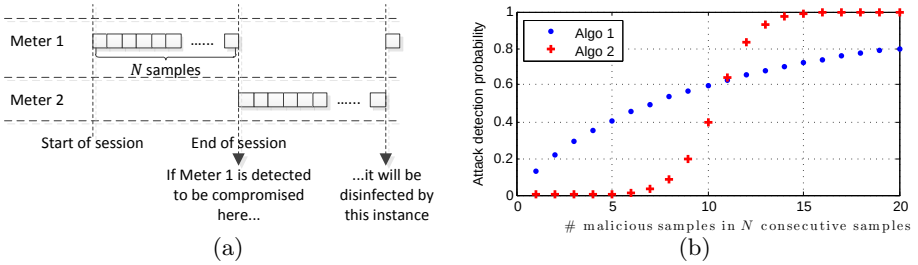


Fig. 6. (a) A session/stage in our security game. (b) Plot of (7) and (8) given $N = 20$, $\alpha_1 = 0.2$, $\beta_1 = 0.8127$, $\alpha_2 = 20$, $\beta_2 = 0.5203$.

d_2 Upon collecting N samples, run Detection Algorithm 2 (to be defined later). If detection result is positive, disinfect the meter.

Detection Algorithms 1 and 2 are *hypothetical* algorithms with attack detection probabilities (true positive rates) of

$$1 - \alpha_1 (x/N)^{\beta_1}, \tag{7}$$

$$\text{and } 1/[1 + e^{-\alpha_2 (x/N - \beta_2)}] \tag{8}$$

where x is the number of malicious samples among N samples; α_1, β_1, α_2 and β_2 are constants. Fig. 6(b) plots the detection probabilities for a set of sample parameters. These definitions are contrived so that Detection Algorithm 1 emulates a clustering-based anomaly detection algorithm, whereas Detection Algorithm 2 emulates a threshold-based algorithm. Detection Algorithm 1 is good for low concentration of malicious samples, while Detection Algorithm 2 is good for high concentration of malicious samples. *It is assumed that the defender can only run one Detection Algorithm at the end of a session due to time constraint.* We emphasize that although we consider two attack actions and two defense actions for numerical simplicity, our approach can be applied to any finite number of attack and defense actions.

The purpose of simulations is to get the state transition matrix $\boldsymbol{M}(a,d) = [M_{s_i,s_j}(a,d)]_{N_S \times N_S}$, and the game matrix $\boldsymbol{G}(s) = [G_{a,d}(s(t))]_{N_A \times N_D}$. $M_{s_i,s_j}(a,d)$ is readily obtained by fixing attacker action at a, defender action at d, and measuring the frequency of encountering states s_i and s_j at the beginning and end of each session respectively. By our definition of risk, $\boldsymbol{G}(s_0) = \boldsymbol{0}$. To obtain $G_{a,d}(s_1)$, we fix attacker action at a, defender action at d, and measure the total load shed during the combined duration of s_1. Suppose the total energy shed is E_{s_1} and the combined duration of s_1 is T_{s_1}, then $G_{a,d}(s_1) = E_{s_1}/T_{s_1}$. In other words, $\boldsymbol{G}(s_1)$ represents the average power shed in state s_1.

To simulate the above security game, we use the system parameters in Fig. 2. Since AGC signals are transmitted to the generating plant once every 2 to 4 seconds [18], we set the sampling rate of the "Defender" and "Attacker" blocks to 2 seconds. Attacks are simulated to start at time 100 s. We set $N = 20$, i.e., 20 samples are read from a meter in each session. The parameters of the Detection

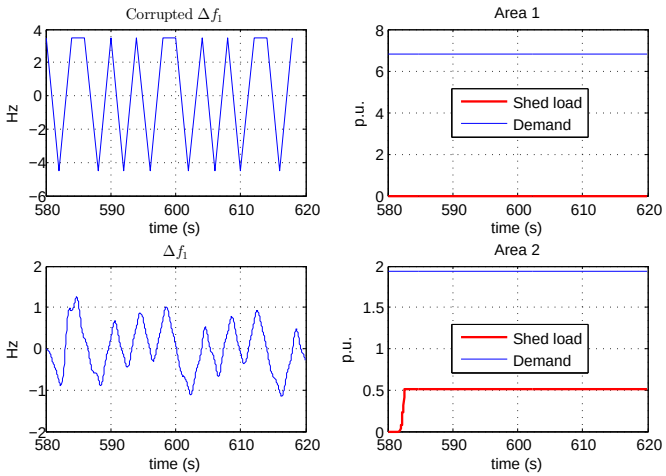


Fig. 7. A sample simulation snapshot spanning two sessions (20 samples per session, 2 s per sample) from time 580 s to 620 s, when attacker action and defender action are fixed at a_2 and d_2 respectively. From 580 s to 600 s, the system consumes false samples from compromised Meter 1, and transitions from state s_0 to state s_1. From 600 s to 620 s, the system consumes false samples from compromised Meter 2, and stays in state s_1.

Algorithms are set according to Fig. 6(b). After a meter is detected to be compromised and disinfected, it will become compromised again after some time; Meter 1 and Meter 2 take 4 sessions and 20 sessions to compromise respectively. Using MATLAB/Simulink, each simulation is conducted for 30 virtual minutes. Fig. 7 shows a simulation snapshot spanning two sessions. The obtained $\boldsymbol{M}$ and $\boldsymbol{G}$ are fed into Algorithm 2. Fig. 8 shows the simulation results, from which the following can be observed:

Effect of sampling rate Since AGC signals are usually transmitted to the generating plant once every 2 to 4 seconds [18], we initially set the AGC sampling rate to 0.5 Hz. A lower sampling rate means a malicious sample will have longer effect on the controller, so when we increase the AGC sampling rate to 1 Hz, the amount of load shed drops conspicuously, as evidenced by the lower-valued $\boldsymbol{G}(s_1)$ (less gain for the attacker). Thus, besides improving control precision, a sufficiently high sampling rate provides a good buffer against attacks. Fig. 8(f, g, h) shows that except for low discount factors, increasing the sampling rate (diminishing the attacker's gain) tend to drive both attacker and defender to adopt a mixed strategy.

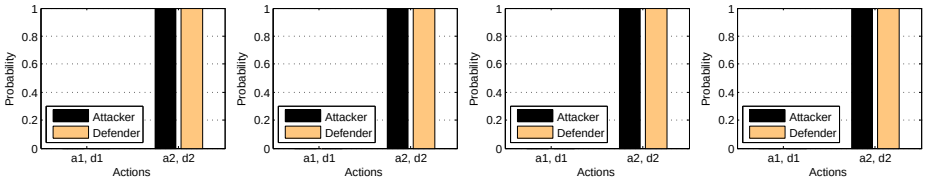
Effect of the discount factor The discount factor α is a logical construct for de-emphasizing the payoff of elapsed stages; it is also a mathematical construct for ensuring convergence. Fig. 8(f, g, h) shows that at a higher sampling rate, varying the discount factor has more impact on defender strategy than on at-

AGC sampling rate: 0.5 Hz

$$\boldsymbol{M}(a_1, d_1) = \begin{bmatrix} 7/11 & 4/11 \\ 4/31 & 27/31 \end{bmatrix} \qquad \boldsymbol{M}(a_1, d_2) = \begin{bmatrix} 9/14 & 5/14 \\ 4/28 & 24/28 \end{bmatrix}$$

$$\boldsymbol{M}(a_2, d_1) = \begin{bmatrix} 8/12 & 4/12 \\ 3/30 & 27/30 \end{bmatrix} \qquad \boldsymbol{M}(a_2, d_2) = \begin{bmatrix} 7/10 & 3/10 \\ 3/32 & 29/32 \end{bmatrix}$$

$$\boldsymbol{G}(s_0) = \mathbf{0} \qquad \boldsymbol{G}(s_1) = \begin{bmatrix} 0.5038 & 0.5884 \\ 0.6643 & 0.6450 \end{bmatrix}$$



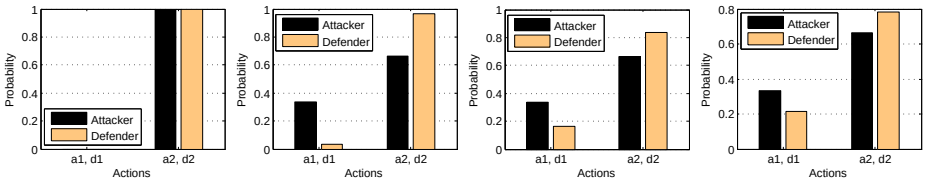
(a) 0.5 Hz, $\alpha = 0.1$ (b) 0.5 Hz, $\alpha = 0.3$ (c) 0.5 Hz, $\alpha = 0.7$ (d) 0.5 Hz, $\alpha = 0.9$

AGC sampling rate: 1 Hz

$$\boldsymbol{M}(a_1, d_1) = \begin{bmatrix} 13/21 & 8/21 \\ 7/64 & 57/64 \end{bmatrix} \qquad \boldsymbol{M}(a_1, d_2) = \begin{bmatrix} 3/8 & 5/8 \\ 4/77 & 73/77 \end{bmatrix}$$

$$\boldsymbol{M}(a_2, d_1) = \begin{bmatrix} 3/9 & 6/9 \\ 6/76 & 70/76 \end{bmatrix} \qquad \boldsymbol{M}(a_2, d_2) = \begin{bmatrix} 5/11 & 6/11 \\ 8/74 & 66/74 \end{bmatrix}$$

$$\boldsymbol{G}(s_0) = \mathbf{0} \qquad \boldsymbol{G}(s_1) = \begin{bmatrix} 0.3046 & 0.3473 \\ 0.3719 & 0.3505 \end{bmatrix}$$



(e) 1 Hz, $\alpha = 0.1$ (f) 1 Hz, $\alpha = 0.3$ (g) 1 Hz, $\alpha = 0.7$ (h) 1 Hz, $\alpha = 0.9$

Fig. 8. Attack and defense strategies organized according to AGC sampling rate and discount factor α.

tacker strategy; and the higher the discount factor, the more often the defender is driven to use action d_1 instead of d_2.

7 Related work

Smart grid cyber security is an emerging area. A comprehensive summary of the challenges confronting this area is provided by Wei et al. [36]: (i) automation components run communication protocols and proprietary operating systems that are designed for connectivity/monitoring/control functionality and not se-

curity; (ii) automation components have limited computational resources due to manufacturing costs and the fact they are used over a long of time exacerbates these resource constraints over time; (iv) resource utilization for performance conflicts with resource utilization for more security.

Substantial research effort is still being dedicated to exploring cyber attacks and their effects on power grids. Stamp et al. [31] develop a cyber-to-physical modeling approach called *Reliability Impacts from Cyber Attack*, for quantifying the degradation of system reliability for a given probability of cyber attack. Several metrics are investigated, including frequency of interruption, loss of load expectancy, load curtailed per interruption, etc. Kundur et al. [16, 17] present two simulation studies – one on a single-generator system, and another on the IEEE 13-bus test system. The studies focus on the effects of attacks by injecting *three* levels of errors into a *single* sensor in the systems. Esfahani et al. [10, 11] design elaborate schemes for controlling maliciously injected AGC output signal to maximally disrupt a grid. Our focus on AGC is in a way inspired by their work. Injecting an AGC output signal potentially requires the attacker to masquerade as an automatic generation controller to a turbine governor, whereas injecting an AGC input signal requires masquerading as a meter to an automatic generation controller. So instead of the AGC output signal, we focus on one of the AGC input signals (i.e., frequency deviation) because from an attacker's perspective, compromising a meter is potentially lower-cost than compromising an automatic generation controller.

Risk assessment has been garnering a lot of attention lately. We note that some authors erroneously refer to risk assessment synonymously as *vulnerability assessment*, which is a different concept [27]. *Attack trees* or attack graphs is a common starting point for most work in this area. An attack tree represents attacks against a system in a tree structure, with the goal as the root node and different ways of achieving that goal as leaf nodes. Cheminod et al. [8] develop a software tool for generating specialized attack trees called *attack and fault propagation graphs*. Ten et al. [34] propose a framework based on attack trees for evaluating system security. They focus on attacks originating from substations connecting to the control center through a VPN. They limit cyber intrusions to firewall penetration and password cracking, singling out password policies and port auditing as the two most important security measures – these assumptions are used in other work by the same research team [30, 33]. Their framework define three vulnerability indices: the *system vulnerability index* is the maximum of *scenario vulnerability indices*, which are products of *leaf vulnerability indices*, which in turn depend on subjective definitions of port vulnerability and password strength. Liu et al. [20] take an attack tree as input, and assign a "difficulty level" to each action on the tree using Analytic Hierarchy Process. Their methodology produces a *vulnerability factor*, an artificial measure of the success probability of an attack. Liu et al. [21] also use Analytic Hierarchy Process–in their case–for assigning weights to performance and security criteria (e.g., "packets burst in local network"). Analytic Hierarchy Process is a decision making methodology that is often applied to risk management, but for its reliance on subjective scor-

ing and failure to satisfy several statistical axioms (e.g., transitivity), the risk management community is divided regarding its validity [14]. In comparison, our work uses only empirical evidence.

The limitation of attack trees is not unrecognized. Sommestad et al. [29] propose *defense graphs* as an alternative to attack graphs, to take into account the countermeasures already in place within a system. They model defense graphs using *influence diagrams*, which are essentially Bayesian networks enhanced with indicators that express *beliefs* on *likelihood* values. The output of their assessment methodology is the expected loss associated with a successful attack. Hahn et al. [13] propose *privilege graphs* to model the privilege states in a system and the paths exploitable by an attacker. The essence of their proposal is an algorithm for computing an *exposure metric*, that takes into account (i) the number of attack paths through the security mechanisms protecting a target asset, and (ii) the path length representing the effort required to exploit a path.

Ten et al. [33] model attacks using *stochastic Petri Nets*, which encapsulate the probability and risk of attacks. They define the metric *system vulnerability* which is the maximum of all *scenario vulnerability* values, and the metric *impact factor* w.r.t to a substation disconnected by a successful attack. Sridhar et al. [30] use stochastic Petri Nets to model computers, firewalls and intrusion protection systems. To assess the *steady-state impact* of attacks on the power system itself, they present the impact study of six coordinated attack scenarios, where coordination is in the sense of targeting multiple power system components at the same time. They define risk as the product of the probability of a successful attack and the resultant shed load; we adopted this definition of risk. Their observation that directly tripping a generator does not always cause more damage than tripping a line coincides with Wang et al.'s [35]. With the exception of [30], most risk assessment work discussed so far is ICT-centric, and does not consider the impact of cyber attacks on the power system itself. In comparison, our work involves the detailed modeling and simulation of attacks on the AGC system.

8 Conclusion and future work

Risk assessment for power grids has been identified as a critical area by the public sector, industry and academia. However, existing risk management standards such as ISO 31000:2009 are more about general principles and guidelines than concrete mathematical techniques. In this work, we identify and assess the risks faced by a critical power system component called automatic generation control (AGC). Our discussion of potential attacks and countermeasures is based on an explicit security threat model. We propose a quantitative risk model capturing the probability and magnitude of security threats faced by the AGC system due to false data injection attacks. Building upon the risk analysis, we model attacker-defender interactions using stochastic (Markov) security games to analyze the best defensive actions under resource constraints. The developed framework is illustrated with a detailed AGC model and simulation results.

For future work, we plan to use more precise models for AGC, turbine governor, generator and underfrequency load shedding. For the most representative models, industrial input is required. In this work, generators are *per convention* simulated as a lumped "System inertia" block, but fine-grained simulations of the electrical circuits in each control area, including the effects of generator tripping triggered by overfrequency protection and islanding, are desirable. In our preliminary study, we consider only attacks on the frequency input to AGC, and only what we call overcompensation attacks among this class of attacks. In future work, we will consider attacks on the tie-line power input, and AGC output. The challenge is to represent these attacks with meaningful attack actions. Economic dispatch is the process of determining how much power each generator generates, and how the power is transmitted under power flow constraints. Since AGC plays a role in economic dispatch, financial loss as a result of attacks interfering with economic dispatch will substantially influence the formulation of the game matrix $\boldsymbol{G}$. We will also take into account communication artefacts such as latency, both as natural occurrences and consequences of attacks.

References

1. Alpcan, T., Başar, T.: Network Security: A Decision and Game Theoretic Approach. Cambridge University Press (2011), `http://www.tansu.alpcan.org/book.php`
2. Alpcan, T., Bambos, N.: Modeling dependencies in security risk management. In: 2009 Fourth International Conference on Risks and Security of Internet and Systems (CRiSIS). pp. 113–116 (Oct 2009)
3. Anderson, G.: Dynamics and control of electric power systems. lecture notes 227-0528-00, ETH Zürich (Feb 2010)
4. Australian Government: Critical infrastructure resilience strategy. ISBN 978-1-921725-25-8, `http://www.tisn.gov.au/` (2010)
5. Baker, S., Filipiak, N., Timlin, K.: In the dark: Crucial industries confront cyberattacks. McAfee 2nd annual critical infrastructure protection report, written with Center for Strategic and International Studies (2011)
6. Bevrani, H.: Robust Power System Frequency Control. Power Electronics and Power Systems, Springer Science+Business Media LLC (2009)
7. Bommannavar, P., Alpcan, T., Bambos, N.: Security risk management via dynamic games with learning. In: Communications (ICC), 2011 IEEE International Conference on. pp. 1–6 (Jun 2011)
8. Cheminod, M., Bertolotti, I., Durante, L., Maggi, P., Pozza, D., Sisto, R., Valenzano, A.: Detecting chains of vulnerabilities in industrial networks. IEEE Transactions on Industrial Informatics 5(2), 181–193 (May 2009)
9. Dzung, D., Naedele, M., Hoff, T.V., Crevatin, M.: Security for industrial communication systems. Proceedings of the IEEE 93(6), 1152–1177 (Jun 2005)
10. Esfahani, P.M., Vrakopoulou, M., Margellos, K., Lygeros, J., Andersson, G.: A Robust Policy for Automatic Generation Control Cyber Attack in Two Area Power Network. In: IEEE Conference on Decision and Control (Dec 2010)
11. Esfahani, P.M., Vrakopoulou, M., Margellos, K., Lygeros, J., Andersson, G.: Cyber Attack in a Two-Area Power System: Impact Identification using Reachability. In: American Control Conference. Baltimore, MD, USA (Jun 2010)

12. Goodin, D.: Rise of "forever day" bugs in industrial systems threatens critical infrastructure. `http://arst.ch/t9d` (2012)
13. Hahn, A., Govindarasu, M.: Cyber attack exposure evaluation framework for the smart grid. IEEE Transactions on Smart Grid 2(4), 835–843 (Dec 2011)
14. Hubbard, D.W.: The Failure of Risk Management: Why It's Broken and How to Fix It. Wiley (2009)
15. Ibraheem, Kumar, P., Kothari, D.: Recent philosophies of automatic generation control strategies in power systems. IEEE Transactions on Power Systems 20(1), 346–357 (Feb 2005)
16. Kundur, D., Feng, X., Mashayekh, S., Liu, S., Zourntos, T., Butler-Purry, K.L.: Towards modelling the impact of cyber attacks on a smart grid. International Journal of Security and Networks 6(1/2011), 2–13 (2011)
17. Kundur, D., Feng, X., Liu, S., Zourntos, T., Butler-Purry, K.: Towards a framework for cyber attack impact analysis of the electric smart grid. In: 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm). pp. 244–249 (Oct 2010)
18. Kundur, P.: Power System Stability and Control. McGraw-Hill Professional (1994)
19. Leitch, M.: Iso 31000:2009—the new international standard on risk management. Risk Analysis 30(6), 887–892 (2010)
20. Liu, N., Zhang, J., Zhang, H., Liu, W.: Security Assessment for Communication Networks of Power Control Systems Using Attack Graph and MCDM. IEEE Transactions on Power Delivery 25(3), 1492–1500 (Jul 2010)
21. Liu, W.X., Fan, Y.F., Zhang, L.X., Zhang, X., Que, H.K.: WAMS information security assessment based on evidence theory. In: International Conference on Sustainable Power Generation and Supply (SUPERGEN '09). pp. 1–5 (Apr 2009)
22. Luo, C., Far, H., Banakar, H., Keung, P.K., Ooi, B.T.: Estimation of wind penetration as limited by frequency deviation. IEEE Transactions on Energy Conversion 22(3), 783–791 (Sep 2007)
23. Machowski, J., Bialek, J.W., Bumby, J.R.: Power System Dynamics: Stability and Control. John Wiley and Sons, Ltd, 2nd edn. (2008)
24. Mounzer, J., Alpcan, T., Bambos, N.: Dynamic Control and Mitigation of Interdependent IT Security Risks. In: 2010 IEEE International Conference on Communications (ICC). pp. 1–6 (May 2010)
25. Mullen, S.K.: Plug-In Hybrid Electric Vehicles as a Source of Distributed Frequency Regulation. Ph.D. thesis, University of Minnesota (2009)
26. NIST: Guidelines for smart grid cyber security. IR 7628 (Aug 2010)
27. NIST: Glossary of key information security terms. IR 7298 Revision 1 (Feb 2011)
28. Prince, B.: Industrial Control Systems are 10 Years Behind Enterprise IT on Security, Say Experts. SecurityWeek.com (Nov 2011)
29. Sommestad, T., Ekstedt, M., Nordstrom, L.: Modeling security of power communication systems using defense graphs and influence diagrams. IEEE Transactions on Power Delivery 24(4), 1801–1808 (Oct 2009)
30. Sridhar, S., Govindarasu, M., Liu, C.C.: Risk analysis of coordinated cyber attacks on power grid. In: Control and Optimization Methods for Electric Smart Grids, Power Electronics and Power Systems, vol. 3, pp. 275–294. Springer US (2012)
31. Stamp, J., McIntyre, A., Ricardson, B.: Reliability impacts from cyber attack on electric power systems. In: IEEE/PES Power Systems Conference and Exposition (PSCE '09). pp. 1–8 (Mar 2009)
32. System, N.C.: Supervisory Control and Data Acquisition (SCADA) Systems. Technical Information Bulletin 04-1 (Oct 2004)

33. Ten, C.W., Liu, C.C., Manimaran, G.: Vulnerability Assessment of Cybersecurity for SCADA Systems. IEEE Trans. Power Syst. 23(4), 1836–1846 (Nov 2008)
34. Ten, C.W., Manimaran, G., Liu, C.C.: Cybersecurity for critical infrastructures: Attack and defense modeling. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 40(4), 853–865 (Jul 2010)
35. Wang, W., Cai, Q., Sun, Y., He, H.: Risk-Aware Attacks and Catastrophic Cascading Failures in U.S. Power Grid. In: 2011 IEEE Global Telecommunications Conference (GLOBECOM 2011). pp. 1–6 (Dec 2011)
36. Wei, D., Lu, Y., Jafari, M., Skare, P., Rohde, K.: An integrated security system of protecting Smart Grid against cyber attacks. In: Innovative Smart Grid Technologies (ISGT). pp. 1–7 (Jan 2010)
37. Wiles, J., Claypoole, T., Henry, P.A., Drake, P., Lowther, S.: Techno Security's Guide to Securing SCADA: A Comprehensive Handbook On Protecting The Critical Infrastructure. Syngress (2008)
38. Wu, F., Moslehi, K., Bose, A.: Power system control centers: Past, present, and future. Proceedings of the IEEE 93(11), 1890–1908 (Nov 2005)