Games and Risk Minimization for Automatic
n Control in Smart Grid

South Australia

S **1,826** CITATIONS

Tansu Alpcan

University of Melbourne

209 PUBLICATIONS   **5,226** CITATIONS

laniswami

Melbourne

NS   **16,652** CITATIONS

Subhrakanti Dey

Uppsala University

90 PUBLICATIONS   **945** CITATIONS

f this publication are also working on these related projects:

g new fetal EKG and fetal doppler signal processing framework for monitoring fetal wellbeing View

Cluster Analysis View project

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

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

Department 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 increasing 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" [9]. **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" [19]. 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 [3]. However, the standard ISO 31000:2009 is "not mathematically based", and has "little to say about probability, data, and models" [13].

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 [5, 17], this work applies game theory to the modeling of attacks on and defenses for a critical power system component called **automatic generation control (AGC)**.

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, due to false data injection attacks.
– A discussion of the security threat model, potential attacks, and countermeasures.
– 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 discusses related work. Section 3 states the problem of assessing the cyber security risks of automatic generation control, an essential power system component. Section 4 presents our game and risk model. In Section 5, we specify an informal threat model; we also discuss attack and defense actions under this threat model. In Section 6, we apply the game and risk model to automatic generation control, and present our simulation results. Section 7 concludes this paper.

2   Related Work

Smart grid cyber security is an emerging area. Substantial research effort is still being dedicated to exploring cyber attacks and their effects on power grids.

Stamp et al. [22] 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. [10] present two simulation studies on the effects of attacks against a single-generator system and a 13-bus system by injecting *three* levels of errors into a *single* sensor in the systems. Esfahani et al. [6, 7] 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. However, we focus on one of the AGC *input signals* (i.e., frequency deviation), since from an attacker's perspective, compromising a meter potentially costs less 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 as *vulnerability assessment*, which is a different concept [19]. *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. Ten et al. [24] 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 *virtual private network*. 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 [21, 23]. 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. [14] 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. Analytic Hierarchy Process is a decision making methodology that is often applied to risk management, but for its reliance on subjective scoring and failure to satisfy several statistical axioms (e.g., transitivity), the risk management community is divided regarding its validity [9]. In comparison, only empirical evidence is used in this work.

The limitation of attack trees is not unrecognized. Sommestad et al. [20] 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. [8] 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. [23] 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. [21] 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. With the exception of [21], 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.

3 Automatic Generation Control in Power Grid

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 [4]. 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) [11, 26]. However, AGC is sometimes referred to as automated (vs manual) load-frequency control [2], or even the entire frequency control system itself [16]. 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 [6]; because of this, it is subject to attacks propagated through the IT system. A detailed threat model is given in Section 5.

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, over-frequency 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 [15, 16], 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 [18]. The gist of the scheme is, when the system frequency drops by more than 0.35 Hz below the nominal frequency, to shed this amount of load:

$$\Delta P_m - \Delta P_e - 0.3/R,$$

where ΔP_m is the change in generator's mechanical power, ΔP_e is the change in generator's electrical power, and R is the droop characteristic. Our aim is to model and quantify the risks posed by an attacker whose intention is to inflict revenue loss on the electricity provider by injecting false data to the automatic generation controller in the hope of triggering load shedding.

For this work, we use the two-area AGC system model and associated simulation parameters in Fig. 1. The automatic generation controller is an integral controller of gain K_{AGC}. We note that AGC design is an established area with designs dating back to the 1950s; a simple integral controller seems to be a logical starting point. The UFLS relay in each area decides on the necessity to shed load, and the amount of load to shed if necessary, using Mullen's algorithm [18]. 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.

In this sample configuration, the maximum sheddable loads are capped at 4 p.u. and 1 p.u. for the 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.
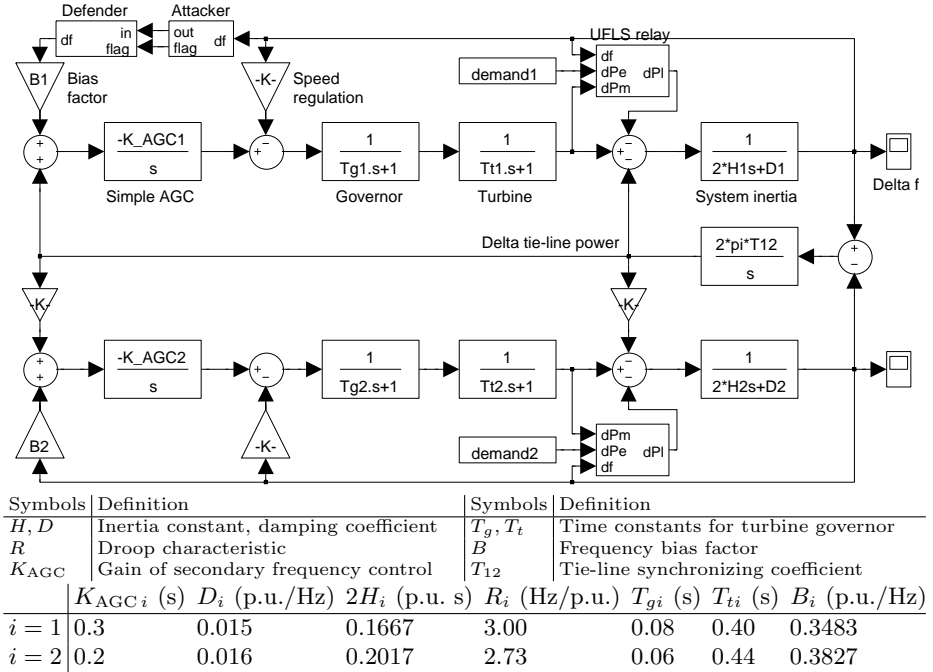
Symbols	Definition		Symbols	Definition
H, D	Inertia constant, damping coefficient		T_g, T_t	Time constants for turbine governor
R	Droop characteristic		B	Frequency bias factor
K_{AGC}	Gain of secondary frequency control		T_{12}	Tie-line synchronizing coefficient

	$K_{\mathrm{AGC}\,i}$ (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. 1. Simulink representation and simulation parameters for a two-area AGC system model based on Bevrani's [4, Fig. 2.10 and Table 2.2]. The top area is labeled area 1. 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. Nominal frequency = 60 Hz.

4   Security Game Model

Our security game model is based on Alpcan and Başar's framework [1]. The concept of **risk states** is combined with 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 (in the unit of power).* Clearly under this definition, risk ranges from 0 to the maximum sheddable load for all areas combined. As a starting point, 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 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 $\mathcal{A}^A \overset{\text{def}}{=} \{a_1, \ldots, a_{N_A}\}$, the defender's action space be $\mathcal{A}^D \overset{\text{def}}{=} \{d_1, \ldots, d_{N_D}\}$,

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 $\boldsymbol{p}^S(t)$ be the probability distribution on the state space $\mathcal{S}$, i.e.,

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

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. As a simplifying assumption, actions have no cost other than their "contribution" to load shedding, so $G_{a,d}(s)$ is the *expected total load shed* in state s under attacker action a and defender action d. In particular, $\boldsymbol{G}(s_0) = \boldsymbol{0}$. Due to the adopted zero-sum Markov game formulation, 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 its own expected total 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. The defender's $\bar{Q}$ at the end of a game is the sum of all realized stage costs discounted by a scalar *discount factor* $\gamma \in [0, 1)$:

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

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 discount factor γ is a logical construct for de-emphasizing the payoff of future stages (smaller $\gamma \implies$ smaller future payoffs). 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). The optimal strategy can be mixed, i.e., stochastic for each state s. At a given stage t, the optimal cost $Q_t(a, d, s)$ (the dependency of s, a and d on t is omitted for notational brevity) can be computed iteratively using the dynamic programming recursion

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

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:

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

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

We can formulate (4) as a linear program:

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

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 (6) and (5) to find V^* and Q^*, is given in Algorithm 2.

Algorithm 2: The value iteration algorithm

Given arbitrary Q_0 and V
repeat
    for $a \in \mathcal{A}^A$ and $d \in \mathcal{A}^D$ **do**
        Update V and Q according to (6) and (5)
    end for
until $V(s) \to V^*$, i.e., $V(s)$ converges

---

5 Threat Analysis

Fig. 2 shows the communication architecture involving a control center and a substation based on the international standard IEC 61850 [23, 26]. Access to the control system in either the control center or the substation is enabled through

a virtual private network (VPN). Some authors [24] 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 (see Fig. 2). This strong attacker model is not entirely unrealistic, however, our goal is to investigate the strategy of an attacker that has successfully penetrated the protected network 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 protected network to trigger cascading failures across the power grid. It is therefore conceivable that an energy provider would make protecting 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 deviation* (Δf) data to the AGC software executing on one of the EMS servers, by compromising one of the meters in the substation (see Fig. 2). 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. Then, it is up to the AGC software to detect such attacks.
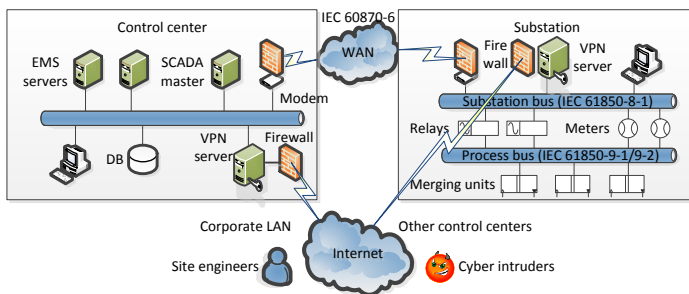

Fig. 2. Accessibility of a power system control center and substation from the Internet. AGC is executed on one of the EMS servers. In our threat model, an attacker can feed the AGC software with false frequency deviation data.

Basic attacks: It is impossible to exhaust all injection patterns, but there are four basic patterns on which more sophisticated attacks are based:

- **Constant injection**: If an attacker injects a constant false Δf, then the it effectively disables the integral control loop, causing the system frequency to converge to a non-nominal frequency. If the false Δf 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.
- **Bias injection**: When the false Δf is a constant bias (displacement) from the true Δf, the effect is similar to that of constant injection because normally the true $\Delta f \approx 0$.
- **Overcompensation**: If the false Δf is k times the true Δf, where k is a large positive number, then the attack effectively causes overcompensation by the integral control loop, and consequently unstable oscillations. As the system frequency sweeps past the overfrequency and underfrequency thresholds, generators will be tripped and loads will be shed, followed by cascading failures. Fig. 3 shows the result of an attack using $k = 8$.
- **Negative compensation**: If the false Δf is $-k$ times the true Δf, where k is a positive number, then the attack effectively reverses the intended effect of the integral control loop, causing the system frequency to diverge from the nominal frequency (see Fig. 3). 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.
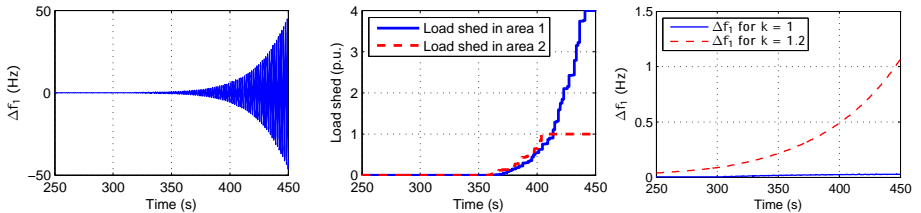

Fig. 3. (Left and Middle) An example of "overcompensation" attack, where the attacker substitutes Δf_1 with $8\Delta f_1$ as frequency input to the area-1 integral controller. As long as the attack persists, neither generator tripping nor load shedding helps stabilize the system. (Right) Negative compensation attack: for large enough k (e.g., 1.2), the system frequency $\to +\infty$.

Basic defenses which are applicable to the overcompensation attack include:

- **Saturation filter**: We can constrain the attack by limiting the Δf 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 [15].

- **Redundancy**: Measurement redundancy is routinely provisioned for critical grid parameters [12]. Multiple frequency meters of different grades can be installed, so that the likelihood of all meters being compromised is small and the AGC software has a non-zero chance of getting genuine frequency data.
- **Detection**: Saturation filtering and redundancy only limit the effect of an attack, stopping an attack requires the attack to be detected and the source be removed. A threshold-based algorithm can be designed to observe the quantity $\sum_t |\Delta f(t)|$; if the quantity is larger than a certain threshold, the system could be under attack. Alternatively, a clustering-based algorithm can be designed to count the number of clusters in the time series $\{\Delta f(t)\}$; if more than one cluster is observed, the system could be under attack.

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 the controller, but on the modeling of system risk dynamics under the actions of the attacker and defender for any given system.

6   Simulation Study

An AGC system under the interactions of an attacker and a defender is simulated in order to observe 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 probability of a session starting in state s_i ends in state s_j. Based on our assumption that actions have no cost other than their "contribution" to load shedding, $\boldsymbol{G}(s_0) = \boldsymbol{0}$; $\boldsymbol{G}(s_1)$ is the expected total load shed in state s_1. 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 $G_{a,d}(s_1) = E_{s_1}/T_{s_1}$.

For numerical simplicity, we define only two attacker actions and two defender actions, although our approach can be applied to any finite number of attacker and defender actions. The chosen attacker actions are:

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 attacker action space $\mathcal{A}^A = \{$ Send N samples, i of which are false $(i = 1, \ldots, N)\}$. The attacker sets a false Δf 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 5.

The defender implements the saturation filter and redundancy measure described in Section 5. For redundancy, the defender 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. 4(a)). Upon collecting N samples, the defender performs one of the following defender actions:

d_1 Run Detection Algorithm 1, a hypothetical algorithm with an attack detection probability of $1 - \alpha_1^{(x/N)^{\beta_1}}$, where x is the number of malicious samples among N samples, α_1 and β_1 are constants. Detection Algorithm 1 emulates a clustering-based anomaly detection algorithm.

d_2 Run Detection Algorithm 2, a hypothetical algorithm with an attack detection probability of $1/[1 + e^{-\alpha_2(x/N - \beta_2)}]$, where x is the number of malicious samples among N samples, α_2 and β_2 are constants. Detection Algorithm 2 emulates a threshold-based algorithm.

We assume that the defender can run only one detection algorithm at the end of each session due to time constraint. If the detection result is positive, the defender disinfects the meter (e.g., by refreshing its firmware, cryptographic keys and so on). Disinfection is assumed to complete within the time frame of one session (see Fig. 4(a)).
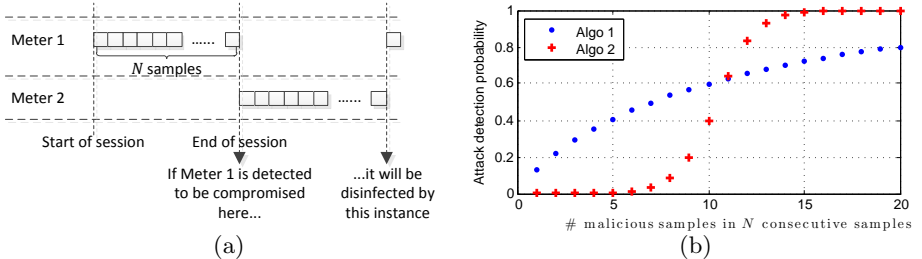

Fig. 4. (a) A session/stage in our security game. (b) Attack detection probabilities of Detection Algorithms 1 and 2 if $N = 20$, $\alpha_1 = 0.2$, $\beta_1 = 0.8127$, $\alpha_2 = 20$, $\beta_2 = 0.5203$.

For simulations, we use the two-area AGC system model and associated simulation parameters in Fig. 1. Since AGC signals are transmitted to the generating plant once every 2 to 4 seconds [11], 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 algorithms are set according to the parameters in Fig. 4(b), such that Detection Algorithm 1 is good for low concentration of malicious samples, while Detection Algorithm 2 is good for high concentration of malicious samples. 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. The obtained M and G are fed into Algorithm 2. Fig. 5 shows the simulation results and the following observations.

Effect of sampling rate: Since AGC signals are usually transmitted to the generating plant once every 2 to 4 seconds [11], 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. When we increase the AGC sampling rate to 1 Hz, the amount of load shed drops conspicuously as evidenced by the lower-valued $G(s_1)$ (less gain for the attacker). Thus, besides improving control precision, a sufficiently high sampling rate provides a good buffer against attacks. Fig. 5(f, g, h) indicates 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: Fig. 5(f, g, h) shows that at a higher sampling rate, where the attacker's gain is lower, defender action d_1 increases in effectiveness as the discount factor increases (future payoffs get more emphasized). In the limit, a pure defense strategy using only d_2 should suffice.

7   Conclusion

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 our preliminary study, we have adopted a risk-neutral framework, such that the expected loss from a blackout tends to conceal the significance of rare events at the tail-end of a probability distribution. Financial risk measures such as value-at-risk and conditional value-at-risk have been proposed to account for these rare events [25], and are being explored in ongoing work. In addition to attacks on the frequency input to AGC, we will consider attacks on the tie-line power input, and the AGC output. We will also use more precise models for AGC, turbine governor, generator and underfrequency load shedding. In this work, generators are *as 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.
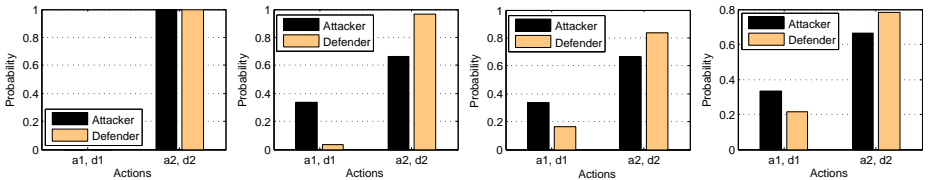
AGC sampling rate: 0.5 Hz

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

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

$$\boldsymbol{G}(s_0) = \boldsymbol{0} \qquad \boldsymbol{G}(s_1) = \begin{bmatrix} 0.5038 & 0.5884 \\ 0.6643 & 0.6450 \end{bmatrix}$$
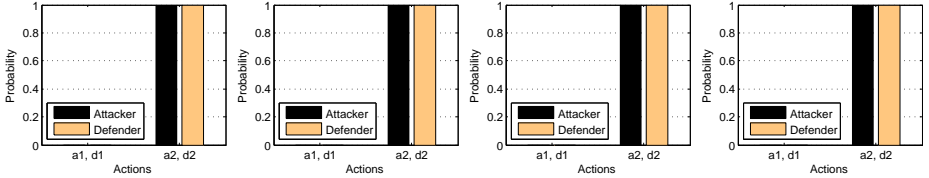
(a) 0.5 Hz, $\gamma = 0.1$ (b) 0.5 Hz, $\gamma = 0.3$ (c) 0.5 Hz, $\gamma = 0.7$ (d) 0.5 Hz, $\gamma = 0.9$

AGC sampling rate: 1 Hz

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

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

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

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

Fig. 5. Attack and defense strategies organized according to AGC sampling rate and discount factor γ.

References

1. Alpcan, T., Başar, T.: Network Security: A Decision and Game Theoretic Approach. Cambridge University Press (2011)
2. Andersson, G.: Dynamics and control of electric power systems. lecture notes 227-0528-00, ETH Zürich (Feb 2010)
3. Australian Government: Critical infrastructure resilience strategy. ISBN 978-1-921725-25-8, http://www.tisn.gov.au/ (2010)
4. Bevrani, H.: Robust Power System Frequency Control. Power Electronics and Power Systems, Springer Science+Business Media LLC (2009)
5. Bommannavar, P., Alpcan, T., Bambos, N.: Security risk management via dynamic games with learning. In: 2011 IEEE International Conference on Communications (ICC). pp. 1–6 (Jun 2011)

6. 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)
7. 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)
8. Hahn, A., Govindarasu, M.: Cyber attack exposure evaluation framework for the smart grid. IEEE Transactions on Smart Grid 2(4), 835–843 (Dec 2011)
9. Hubbard, D.W.: The Failure of Risk Management: Why It's Broken and How to Fix It. Wiley (2009)
10. 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)
11. Kundur, P.: Power System Stability and Control. McGraw-Hill Professional (1994)
12. Lefebvre, D., Bernard, S., Van Cutsem, T.: Undervoltage load shedding scheme for the Hydro-Québec system. In: IEEE Power Engineering Society General Meeting. vol. 2, pp. 1619–1624 (Jun 2004)
13. Leitch, M.: ISO 31000:2009—The New International Standard on Risk Management. Risk Analysis 30(6), 887–892 (2010)
14. 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)
15. 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)
16. Machowski, J., Bialek, J.W., Bumby, J.R.: Power System Dynamics: Stability and Control. John Wiley and Sons, Ltd, 2nd edn. (2008)
17. 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)
18. Mullen, S.K.: Plug-In Hybrid Electric Vehicles as a Source of Distributed Frequency Regulation. Ph.D. thesis, University of Minnesota (2009)
19. NIST: Glossary of key information security terms. IR 7298 Revision 1 (Feb 2011)
20. 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)
21. 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)
22. 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)
23. 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)
24. 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)
25. Varaiya, P., Wu, F., Bialek, J.: Smart operation of smart grid: Risk-limiting dispatch. Proceedings of the IEEE 99(1), 40–57 (Jan 2011)
26. Wu, F., Moslehi, K., Bose, A.: Power system control centers: Past, present, and future. Proceedings of the IEEE 93(11), 1890–1908 (Nov 2005)