Published article:

T. R. B. Kushal, M. S. Illindala and J. Wang, "Analytical Risk Assessment of Communication Cyber Attacks on Automatic Generation Control," *2022 IEEE Industry Applications Society Annual Meeting (IAS)*, Detroit, MI, USA, 2022, pp. 1-8, doi: 10.1109/IAS54023.2022.9939906.

Analytical Risk Assessment of Communication Cyber Attacks on Automatic Generation Control

Tazim Ridwan Billah Kushal
Graduate Student Member, IEEE
The Ohio State University
2015 Neil Ave, Dreese Lab 205
Columbus, OH 43210, USA
kushal.1@osu.edu

Mahesh S. Illindala
Senior Member, IEEE
The Ohio State University
2015 Neil Ave, Dreese Lab 205
Columbus, OH 43210, USA
millindala@ieee.org

Jiankang Wang
Member, IEEE
The Ohio State University
2015 Neil Ave, Dreese Lab 205
Columbus, OH 43210, USA
wang.6536@osu.edu

***Abstract*—Cybersecurity has become a crucial consideration in critical power system applications such as automatic generation control (AGC) due to the increasing use of information and communication technology. The existing research literature includes several descriptions of threats from either physics-based or pure cybersecurity perspectives. A holistic cyber-physical security assessment method is necessary to guide future decisions regarding AGC organization. This paper develops an analytical risk assessment method for integrity attacks on AGC communication while considering the cyber-physical causal chain. Attack occurrences and detections are modeled as stochastic events, while considering their physical impact and modeling accuracy of mitigation measures. The results provide a holistic cyber-physical security assessment and recommendations for securing the AGC system against compromised communications.**

***Index Terms*—Automatic generation control, SCADA, smart grid, wide area networks.**

I. Introduction

Widespread adoption of information and communication technology (ICT) in industrial control systems has transformed the contemporary electric grid is a cyber-physical system (CPS) where the cyber layer (computers, network devices, embedded systems) interacts directly with the physical infrastructure (machines, transmission lines, switchgear). There is increased use of ICT in various grid monitoring and control applications, such as intelligent electronic devices (IEDs) for advanced measurement, protection, and control capabilities [1] and the Internet Protocol (IP) for scalable and robust wide area communication [2]. The cyber-physical paradigm involves novel security threats and cyber attacks on the power grid can have severe consequences, especially because traditional ICT security measures may not be compatible with operational reliability constraints [2], [3] and specialized protocols used for monitoring and control [4].

Load frequency control (LFC) is a critical application that is susceptible to integrity attacks aiming to degrade dynamic performance and cause instability [5]. Secondary LFC, which is part of automatic generation control (AGC) and regulates frequency across multiple areas, provides a higher cyber attack surface due to the necessity of communication over wide area networks (WANs). Several authors have provided physics-based descriptions of attacks on LFC and corresponding detection and mitigation measures [5]–[7]. Limitations of these methods include the ability of a deliberate attacker to bypass known error-checking mechanisms such as bad data detection and Kalman filtering [8], [9], absence of traditional ICT security measures such as intrusion detection systems (IDSs) and failure to account for the cyber mechanisms of attack.

Some authors have approached the topic from a cybersecurity perspective. Process control systems such as AGC tend to have static topology, regular traffic patterns, and simple protocols [10]. This enables specification-based detection [11] that can be augmented by data-driven methods [12]. Many authors treat the detection module as a centralized application that can screen the grid for attacks. Specific location of the detector is considered in some works, such as [13] where the placement is critical for protection systems. However, location of detection modules is relevant to AGC attacks as well due to the existence of network-based attacker vectors such as denial-of-service (DoS) and man-in-the-middle (MITM) attack. Host-based detection may be unable to defend against an attack that blocks or intercepts communication between two hosts. Network-based detection has been proposed as an alternative, as in [14] where IDSs are placed at different layers in the smart grid. Not all attacks generate abnormal traffic patterns detectable by network-based IDS [15], so a hybrid implementation involving host-based detection is desirable.

However, true CPS security assessment requires a holistic approach where cyber events can be related to physical ones in a manner where the causal chain can be explored. Purely physics-based methods often fail to account for factors such as stochasticity and attack vectors, while a heavy focus on the cyber domain usually lacks direct impact analysis of cyber attacks in terms of physical consequences.

Evaluating the efficacy of detection schemes requires risk analysis of cyber attacks. Risk assessment studies are generally application-specific, focusing on certain attack vectors and their impact on the power system. A method for security risk assessment of protection systems is developed in [16]. Game-theoretic stochastic frameworks [17], [18] are used to account for uncertainty and multiple possible strategies. An analytical framework for optimizing defensive strategies and minimizing risk of attacks on AGC was proposed in [17].

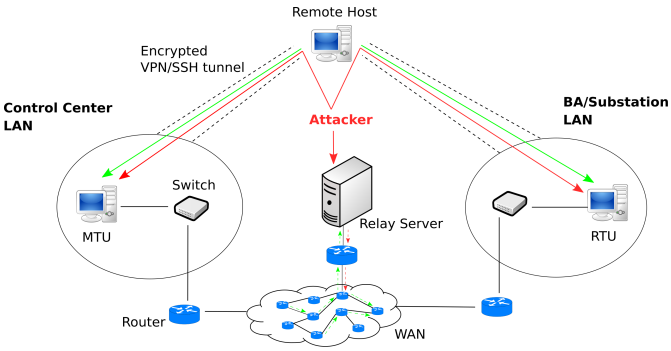

Fig. 1. Remote access points for MITM cyber attacks.

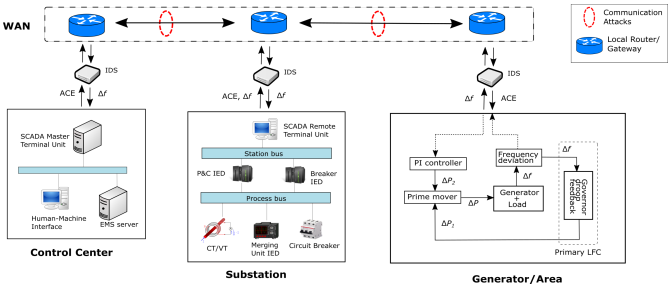

Fig. 2. Overview of WAN communication with potential locations for attack and detectors.

The risk evaluation method in [17] is limited by the use of a stationary stochastic process and the conditional value-at-risk approach based on an empirical loss distribution. Since LFC is a dynamic control application, expected load shedding does not adequately capture the risk of a disruptive attack.

This paper develops holistic cyber-physical security for integrity attacks on AGC communication using analytical risk assessment. The proposed risk assessment methodology derives risk metrics that quantify the potential impact of compromised signals on LFC dynamics. Impact of countermeasures against such attacks also needs to be considered, since AGC is a centralized control application where discarding an untrusted signal has potential consequences. This study considers various types of risk related to integrity attacks and proposes metrics for them. To enable overall interpretation of risk profile, a composite formula combining the different categories is also proposed. Section II describes the high-level model of cyber-physical security and different categories of risk associated with communication attacks. Section III defines quantitative risk metrics and describe their respective effects on system dynamics. Finally, Section IV provides the risk profiles of an AGC testbed using the proposed analytical risk assessment method.

II. Cyber-Physical Security Model

A. Detection and Mitigation System

In an industrial control system such as AGC, both ICT-based and physics-based methods of cyber attack detection and mitigation are applicable. Physics-based methods rely on

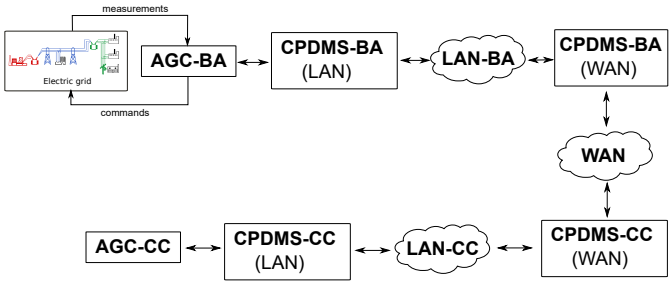

Fig. 3. Overview of the CPDMS countermeasure against communication cyber attacks on AGC, shown in terms of information exchange between a CC and a BA area.

mathematical models of the physical process while ICT-based methods are only concerned with information flows in the cyberspace. Therefore, the two types of methods focus on different aspects of cyber attacks, namely the cyber- and physical-side features. Accurate CPS security assessment requires a holistic approach where cyber events can be related to physical ones in a manner where the causal chain can be explored. Purely physics-based methods often fail to account for factors such as stochasticity and attack vectors, while a heavy focus on the cyber domain usually lacks direct impact analysis of cyber attacks in terms of physical consequences. To overcome their respective shortcomings, both types of methods are combined into a general detection and mitigation system, referred to as a Cyber-Physical Detection and Mitigation System (CPDMS). CPDMS is used as a general umbrella term that, for the sake of holistic security assessment, includes the entire suite of physics- and ICT-based techniques for cyber attack mitigation and detection.

Figure 3 provides an overview of the proposed CPDMS architecture to counter communication cyber attacks on information exchange between CC and BA. The suffixes “BA” and “CC” are used to identify the local components or instances of applications. Implementing CPDMS on both BA and CC sides provides the flexibility to use detection and mitigation on either side when information flow in one direction is compromised. BA and CC components can check incoming signals but not outgoing ones. While CPDMS-BA can screen commands sent by the CC, it is unable to check frequency measurements sent to the CC. Due to the packet encapsulation principle of ICT communication protocols, relevant information that could potentially enable threat detection may be removed when network traffic passes from a WAN to a LAN. Therefore, separate WAN components are specified for CPDMS to check incoming traffic from the WAN.

B. Categories of Risk

In a stochastic model where an intrusion detection system (IDS) attempts to identify compromised signals, there are four possible outcomes: true negative (TN), false positive (FP), true positive (TP), and false negative (FN). A positive identification means that the IDS has classified a signal as being attacked. Detection accuracy is characterized by conditional

probabilities of correctly detecting attacks p_{tp} and mistaking legitimate signals for compromised ones p_{fp}, obtained from receiver operating characteristics (ROC) analysis [19] of the detector. Attack probability p_a is the probability of at least one exposed link being attacked.

An outcome of TN represents normal operation, where no attack occurs and none is detected. The other three outcomes can disrupt AGC operation in different ways and are represented by three different types of risk, as described below:

1) *Undetected Attack Risk* (UAR): When CPDMS fails to detect an attack (ie. an FN outcome), the compromised signal may impact AGC and cause the system frequency to deviate from normal operation. UAR quantifies the risk posed by such undetected attacks.
2) *False Alarm Risk* (FAR): The detector may also mistakenly flag an uncorrupted signal (ie. an FP) as an attack and unnecessarily disrupt frequency control. The impact of such misdetections is measured by FAR.
3) *Blocked Signal Risk* (BSR): Although research literature focusing on ICT-based threats consider risk posed by FNs and FPs, a third type of risk is posed by TPs in a synchronous time-critical application such as AGC. A compromised signal correctly identified by CPDMS cannot be used since the information it contains cannot be trusted. However, if the lost signal cannot be reconstructed with 100% accuracy, this outcome can potentially impact frequency control, as quantified by BSR.

It should be noted that different types of risk should be interpreted differently in terms of impact on system dynamics. UAR represents the risk of system instability (undamped oscillations) while FAR and BSR represent the risk of undesirable frequency response which would result in suboptimal performance. Two scenarios with the same overall risk but different proportions of UAR, FAR, and BSR are not directly comparable. However, scenarios can be directly compared within the same risk category. For example, a scenario with high UAR is much more likely to be unstable than a scenario with low UAR.

III. Analytical Risk Assessment Model

A. Quantitative Risk Metrics

For the purpose of quantitative analysis, this section develops metrics of each type of risk enumerated in Section II-B. The proposed metrics possess the quality of ergodicity, ensuring convergence to a fixed value for a given probability distribution instead of random variations caused by differing outcomes of the stochastic process.

1) Undetected Attack Risk: The general state-space representation of communication-based attacks, shown in Appendix A, is the basis of the UAR metric, which captures the impact of undetected attacks on frequency dynamics. A more detailed derivation is presented in [20].

A system that has eigenvalues with real positive part α_τ will be unstable. In this case, the eigenvalues may change from one interval to another, since α_τ is a function of ω_τ. Therefore, instead of a single eigenvalue, a series of eigenvalues needs to be considered. The formula in 14 suggests that the final state depends on $e^{\left(\sum_{\tau=1}^{T}\alpha_\tau\right)}$. Therefore the sum $\sum_{\tau=1}^{T}\alpha_\tau$, where $\alpha_\tau = \lambda_{\max}(\mathbf{A}_\tau)$ is the largest eigenvalue of $\mathbf{A}_\tau$, represents the risk to the system over T successive periods.

$$\text{UAR} = \lim_{T\to\infty}\frac{1}{T}\sum_{\tau=1}^{T}\alpha_\tau = \sum_{\omega\in\Omega} p_\omega \alpha_\omega \tag{1}$$

Since α_τ is a random variable, its summation over a fixed set of periods does not converge and is therefore not ergodic. However, the time-average of α_τ over T periods does converge to its expectation value, which is the sum of eigenvalues weighted by outcome probabilities, as T approaches infinity. This quantity is used to define UAR as shown in (1), where p_ω and α_ω represent the probability and largest eigenvalue for scenario ω respectively.

2) False Alarm Risk: It is expected that the attacker will modify $\mathbf{A}_\tau$ to ensure that unstable eigenvalues exist for FN outcomes. In case of FPs and TPs, where such eigenvalues are unlikely to exist in a properly designed controller, the UAR metric does not provide a useful measure of risk.

$$\text{FAR} = (1-\theta)\sum_{\omega\in\Omega} p_\omega n_{\text{FP}}^{(\omega)} \tag{2}$$

Instead, FAR is expressed using the formula in (2), which is the expectation value of the number of FPs n_{FP} multiplied by reconstruction error rate. A high reconstruction accuracy θ can reduce FAR by decreasing reliance on ACE commands.

3) Blocked Command Risk: Similar to FAR, BSR is defined using the number of TPs n_{TP} as follows:

$$\text{BSR} = (1-\theta)\sum_{\omega\in\Omega} p_\omega n_{\text{TP}}^{(\omega)} \tag{3}$$

To determine the overall risk profile, UAR, FAR, and BSR must be combined into a single formula, where the different types of risk are assigned their respective weights. The weights are assigned on the basis of how significant each type of risk is, which is determined by the outcome of the stochastic process and its impact of frequency dynamics.

B. Impact of Various Types of Risk

1) Precision and Recall: Significance of outcomes is quantified by the precision ($P = \frac{TP}{TP+FP}$) and recall ($R = \frac{TP}{TP+FN}$) ratios, which measure the ability to avoid false alarms and undetected attacks respectively. Precision and recall depends on the probabilities p_a, p_{tp} and p_{fp}. However, their relative proportions can be used to assign weights.

Table I shows the situations where different types of risk become more significant. With fewer FPs and higher FNs (high precision, low recall), the key threat to the system is posed by UAR. When the reverse is true (low precision, high recall), the large number of FPs drives up the significance of FAR. When both precision and recall have moderate values, which

TABLE I
WEIGHTS ASSIGNED TO RISK CATEGORIES BASED ON PRECISION AND RECALL

Precision (P)	Recall (R)	Weight	Risk
high	low	$\frac{P}{P+R}$	UAR
low	high	$\frac{R}{P+R}$	FAR
medium	medium	$2\frac{PR}{P+R}$	BSR

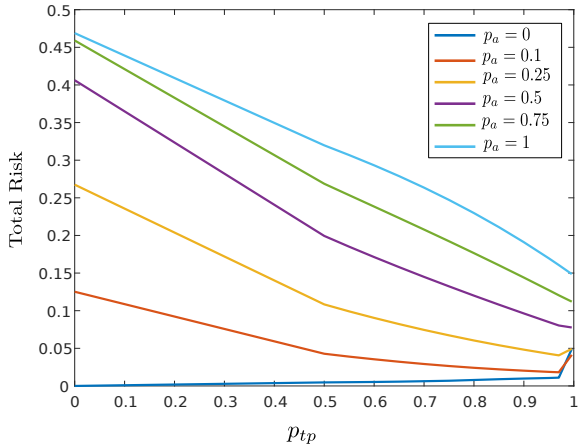

Fig. 4. Total risk plotted against detection accuracy p_{tp} for different values of attack probability p_a.

is desirable in an accurate detection system, BSR becomes the most prominent type of risk. Accordingly, each type of risk is assigned a weight which is maximized when it becomes the most significant contributor to overall risk.

2) Dynamic Impact on Frequency Response: A second factor to consider is the impact of each type of error on frequency dynamics. It is expected that in case of an attack capable of inducing system instability, FNs will have a much larger impact than FPs and TPs. The impulse response of the system in scenario ω is used as a measure of the impact on frequency response.

$$\text{IR}_\omega = |1 - g_\omega h_\omega \theta_\omega| \frac{\beta}{T_{ch} T_g M} \tag{4}$$

The initial non-zero impulse response IR for the state-space model described in Appendix A under outcome ω is given by (4), where g_ω, h_ω and θ_ω are attack and defense parameters from the model described in A-B, specific to the scenario ω. The remaining parameters are from the linearized LFC dynamic model that has been used as the basis of many studies including [5]: β is the frequency bias factor in the AGC loop, and T_g, T_{ch} and M are constants from the governor control loop. A detailed derivation of this expression can be found in [20].

3) System Inertia Characteristics: Aside from stochastic outcomes, the frequency response characteristics of the system also determines the level of risk. Low-inertia systems are more vulnerable than high-inertia ones to the same cyber attack vectors, since the same power demand fluctuations will generally cause larger frequency fluctuations. The concept of measured effective inertia [21] is included in the weights to account for frequency response characteristics. This will allow proper comparison between systems with different levels of inertia.

$$\text{MEI} = \frac{\int_{t_0}^{t} \Delta P_D(t) dt}{\Delta f(t)} \tag{5}$$

Measured effective inertia (MEI) of the system is expressed in (5) as the ratio of power change ΔP_D to frequency change Δf. MEI under normal operation (no attacks and no detection events) is used to linearly scale the overall risk, since it is applicable regardless of the stochastic outcomes. The reciprocal of MEI assigns a higher weight to high-risk low-inertia systems.

4) Composite Formula for Overall Risk:

$$W_{\text{UAR}} = \frac{P}{P+R} \times \text{IR}_{\text{FN}} \times \frac{1}{\text{MEI}} \tag{6}$$

$$W_{\text{FAR}} = \frac{R}{P+R} \times \text{IR}_{\text{FP}} \times \frac{1}{\text{MEI}} \tag{7}$$

$$W_{\text{BSR}} = 2\frac{PR}{P+R} \times \text{IR}_{\text{TP}} \times \frac{1}{\text{MEI}} \tag{8}$$

$$\text{Total Risk} = W_{\text{UAR}}\text{UAR} + W_{\text{FAR}}\text{FAR} + W_{\text{BSR}}\text{BSR} \tag{9}$$

Both IR and MEI are averaged over multiple BA areas. Weights for the three categories of risk are assigned as shown in (6)–(8). The overall risk, given by (9), is calculated as the weighted sum of the different types of risk.

IV. SIMULATION RESULTS AND ANALYSIS

For this study, we are concerned with high-impact attacks that can severely disrupt grid operation. Some attacks have minimal effects on frequency stability and therefore are not considered. From the four basic injection patterns described in [17] by Law et al., the overcompensation and negative compensation patterns are used in this study. In this attack pattern, frequency control is disrupted by modifying the measurements and command by a scaling factor. Overcompensation (positive scaling factor) is found to have relatively small effect on frequency response. Although the effect might be increased by using higher scaling factors, multiple researchers [7], [8], [22] have observed that arbitrarily high modifications of signals raises the likelihood of physics-based detection. Therefore, as commented in [17], it is logical to assume some realistic constraints on the capabilities of an adversary.

Risk assessment metrics defined in Section III are calculated for different values of attack probability, detection accuracy for attacks and estimation accuracy for blocked commands. The quantitative model includes the threat posed by both adversarial actions (allowing compromised commands) and corresponding countermeasures (blocking untrusted commands). The IEEE 39-bus system, divided into 3 balancing authority areas, is used as the AGC test system for this study.

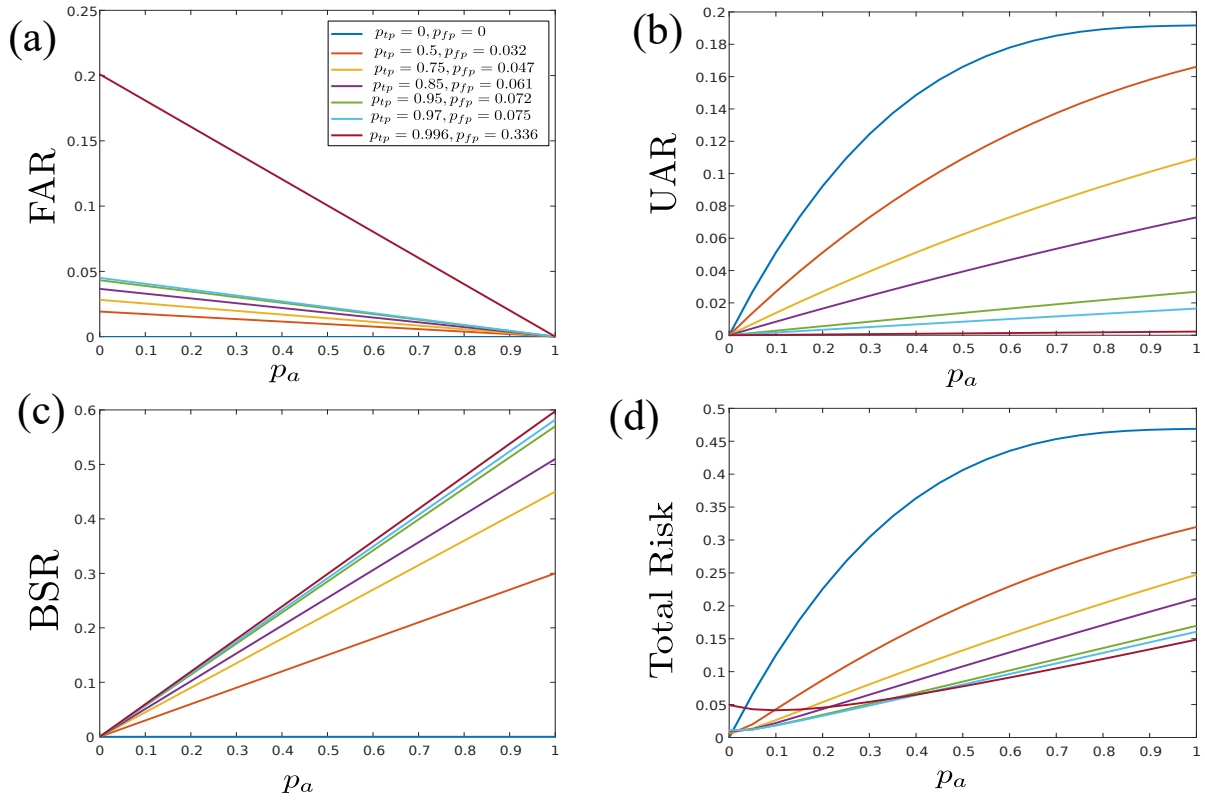

Fig. 5. Risk profile for various categories and overall metric plotted against attack probability p_a for different values of detection accuracy p_{tp}, with fixed estimation accuracy $\theta = 0.8$.

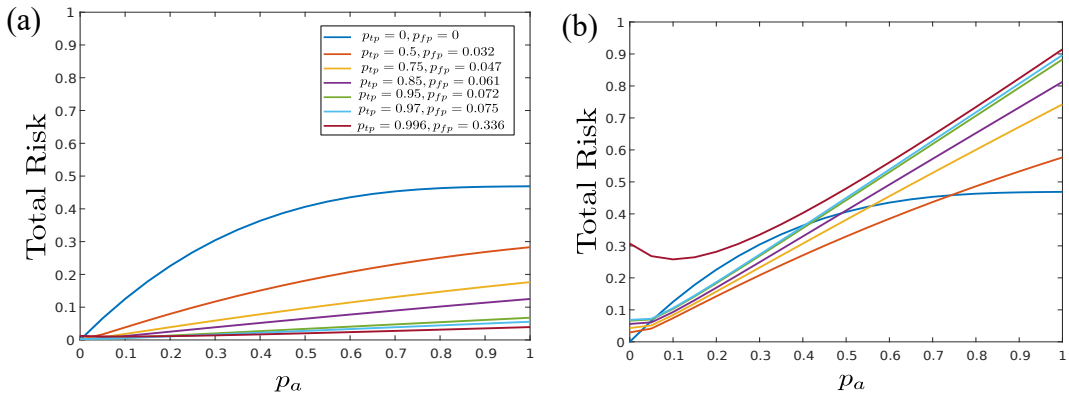

Fig. 6. Total risk to the system with varying p_a for (a) $\theta = 0.9$, and (b) $\theta = 0.5$.

A. Attack Probability and Detection Accuracy

In Figure 5, as the probability of attack increases, UAR and BSR rise while FAR declines. Increasing detection accuracy p_{tp} also raises p_{fp}, leading to increase in both BSR and FAR, although the total risk decreases due to the declining number of FNs. The risk profile in Figure 5(d) indicates that in an extreme case, with a high FP rate (33.6%) and low attack probability ($p_a \approx 0$), the total risk is higher for CPDMS than the system with no countermeasures. However, generally the CPDMS significantly lowers the system risk across the full range of attack frequency values. The same trend is observed when total risk is plotted against p_{tp} in Figure 4. In the vast majority of cases, the risk declines with increasing p_{tp}. Only when the detection accuracy approaches 100% and attack probability is relatively low ($p_a \leq 0.25$) does the risk increase slightly.

B. Estimation Accuracy for Blocked Signals

Risk assessment using the proposed method indicates that higher detection accuracy is generally favored in securing the power grid against attacks on AGC communication. However, the results so far have assumed the estimation accuracy of blocked signals to be fixed at 80%. Setting $\theta = 0.8$ means that

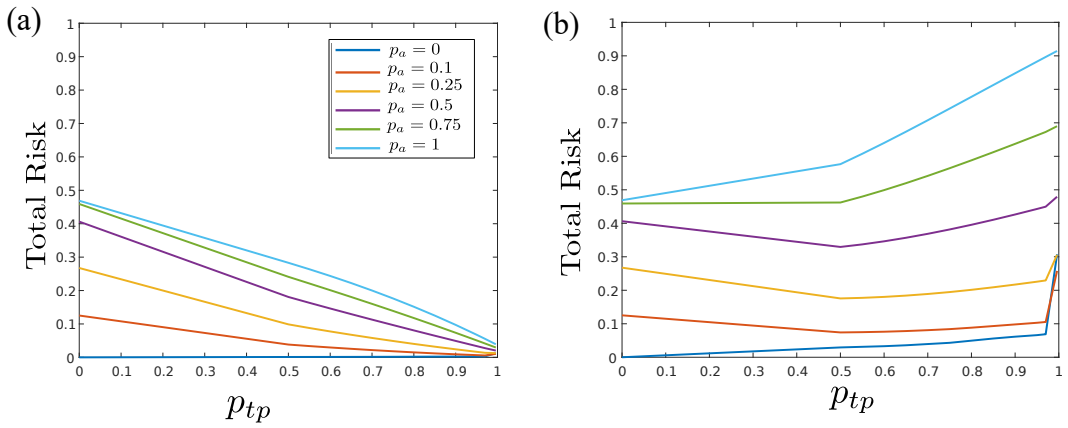

Fig. 7. Total risk to the system with varying p_{tp} for (a) $\theta = 0.9$, and (b) $\theta = 0.5$.

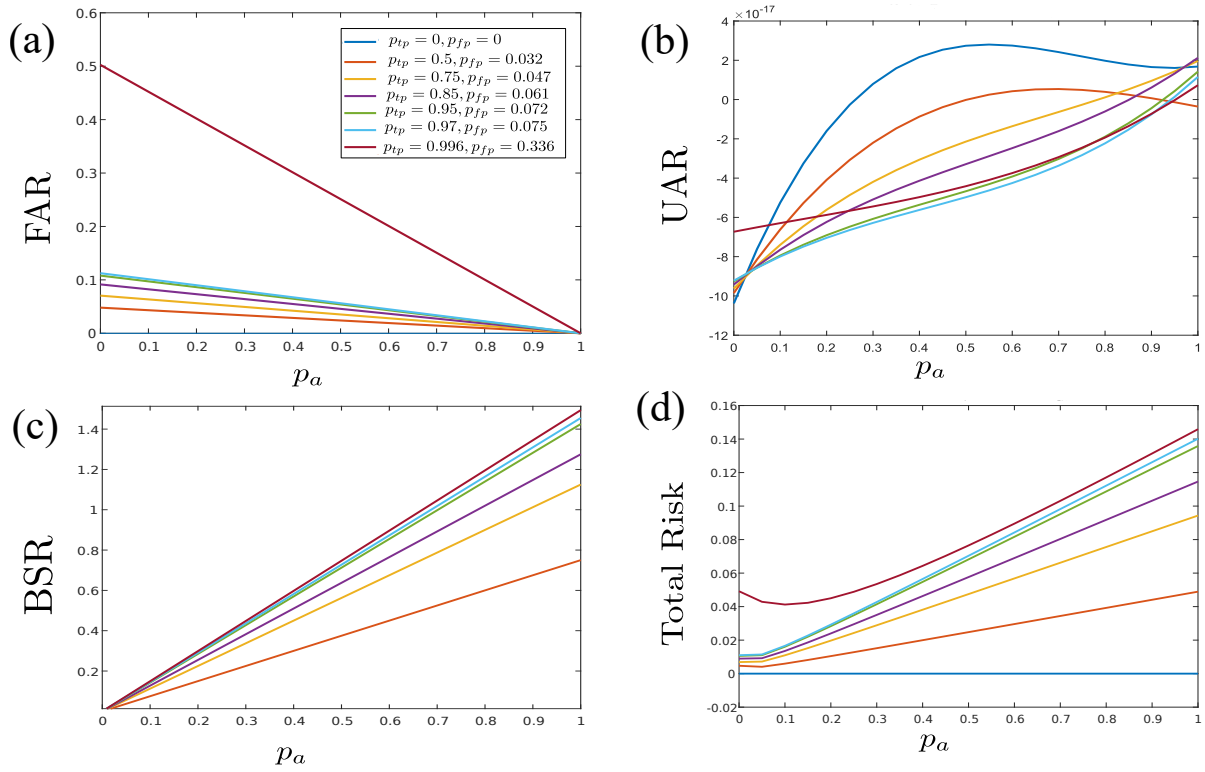

Fig. 8. Risk profile for various categories and overall metric plotted against attack probability p_a for different values of detection accuracy p_{tp}, with fixed estimation accuracy $\theta = 0.8$, when frequency deviation measurements are attacked instead of control commands to the generators.

for each blocked signal (both FPs and TPs) the change in load references setpoints for generators in the affected BA area is 80% of its actual value. Figure 6 demonstrates the significant effect of estimation accuracy on the risk profile when all other parameters are equal. While a high estimation accuracy (90%) causes the familiar pattern of declining risk with high p_{tp}, setting $\theta = 0.5$ results in a more ambiguous pattern where higher p_{tp} does not correspond to lower risk, even at high attack frequencies. Low estimation accuracy results in high FAR at low p_a and high BSR at high p_a, which combine to match or exceed UAR. Reducing the estimation accuracy reverses the pattern of declining risk for rising detection accuracy, as shown in Figure 7, as well as increasing overall risk.

As discussed in Section II-B, actual impact on frequency dynamics depends on the type of risk. Although Figures 6 and 7 appear to indicate that an inaccurate estimator makes low detection accuracy (or even no detection at all) preferable, it should be noted that UAR remains the same in both cases. Low θ causes FAR and BSR to increase, thus increasing total risk. While the risk of instability remains the same, higher p_{tp} leads to greater risk of undesirable deviation from nominal

frequency.

C. Effect of Attack Injection Point

Cyber attacks on AGC communications may target either measurements or commands. In the modified state-space model described in Appendix A, the scaling factors g_i and h_i change the frequency deviation and control input for the i-th BA area, respectively. The results so far have used $h = -1$, since modifying the commands was observed to have greater impact on risk metrics and frequency dynamics. In this section, the alternative method of using $g = -1$ is studied to demonstrate the effect of altering the injection point. In case of corrupted measurements, the detection is assumed to take place on the CC side, so that the centralized CPDMS-CC is used instead of decentralized CPDMS-BA (see Figure 3).

Risk metrics for $g = -1$, shown in Figure 8, are markedly different than the case of $h = -1$ (as shown in Figure 5) for the same values of p_a and $p_t p$. Both FAR and BSR are higher because centralized detection raises both TP and FP rates while reducing FN rate. However, since UAR is close to zero, total risk is much lower. The plots in Figure 8 show that FAR and BSR account for most of the risk profile and frequency instability is unlikely.

D. Incentive for Decentralized Control

The numerical results presented in this section indicate that minimizing risk of instability and frequency fluctuations under varying conditions of attack requires maximization of both detection and estimation accuracy. A highly accurate detection system paired with an inaccurate mitigation system degrades performance during periods of low cyber threat. On the other hand, the absence or low accuracy of detectors exposes the system to potential instability when a non-negligible cyber threat exists. Both kinds of risk can only be minimized by improving the accuracy of detection and mitigation simultaneously. This observation provides an indirect argument for decentralization in power system control applications.

V. Conclusion

Cybersecurity threats have become a prominent consideration for the modern smart grid, which relies increasingly on ICT for many control applications. This study presents a risk assessment method for integrity attacks on AGC, which is a fully automated control application designed to function without human operators. The proposed metrics, based on various types of risk and their respective impacts on frequency dynamics, use analytical formulae derived from probability distributions of adversarial activity and detection accuracy. The focus of this study was the quantification of high-risk factors by developing suitable metrics for different types of risk. Roles played by the various risk categories is demonstrated by the simulation results, which show the effects of changing attack probability, detection accuracy, and estimation accuracy for blocked signals. The results also indicate that the attack causal chain is also an influencing factor, since attacking frequency deviation measurements was observed to pose significantly lower risk of disruption. The study indirectly provides support for increasing decentralization of control, since accurate estimation of blocked or corrupt signals can alleviate risk posed by cyber attacks irrespective of other factors.

Appendix A
State-Space Representation of Communication Attacks

A. Recursive State Transition Function

The mathematical model of linearized LFC dynamics has been used to study cyber attacks and can be found in studies such as [5]. For simplicity, let us consider one possible solution from the fundamental set of solutions and assume $\mathbf{x}_\tau(t) = c_\tau \mathbf{v}_\tau e^{\lambda_\tau t}$ for time interval τ. Since the general solution is a linear combination of such solutions, the results can be generalized. For a complex eigenpair with eigenvalue $\lambda_\tau = \alpha_\tau + \beta i$ and eigenvector $\mathbf{v}_\tau = \mathbf{a}_\tau + \mathbf{b}_\tau \mathbf{i}$, the solution is

$$\begin{aligned}\mathbf{x}_\tau(t) =& c_{1,\tau} e^{\alpha_\tau t}(\mathbf{a}_\tau \cos\beta_\tau t - \mathbf{b}_\tau \sin\beta_\tau t) + \\ & c_{2,\tau} e^{\alpha_\tau t}(\mathbf{a}_\tau \sin\beta_\tau t - \mathbf{b}_\tau \cos\beta_\tau t)\end{aligned} \tag{10}$$

The system can be defined recursively as

$$\mathbf{x}_{\tau+1} = (\mathbf{Q}_{\tau+1} + p_{\tau+1}\mathbf{x}_\tau) e^{\alpha_{\tau+1}\Delta t} \tag{11}$$

$$p_\tau = \cos\beta_\tau \Delta t \tag{12}$$

$$\mathbf{Q}_\tau = (c_{2,\tau}\mathbf{a}_\tau - c_{1,\tau}\mathbf{b}_\tau)\sin\beta_\tau \Delta t \tag{13}$$

where $\mathbf{x}_\tau$ is the state of the system at the end of the τ-th time interval. Let $\mathbf{x}_0$ be the state at the start of the first interval ($\tau = 1$). Then the state after T intervals can be expressed as

$$\begin{aligned}\mathbf{x}_T =& \mathbf{Q}_T e^{\alpha_T \Delta t} + \mathbf{x}_0 \prod_{\tau=1}^{T} p_\tau e^{\alpha_\tau \Delta t} + \\ & \sum_{k=1}^{T-2} \left(\prod_{\tau=T-k+1}^{T} p_\tau e^{\alpha_\tau \Delta t} \right) \mathbf{Q}_{T-k} e^{\alpha_{T-k}\Delta t}\end{aligned} \tag{14}$$

Unlike p_τ, $\mathbf{Q}_\tau$ for a particular interval is not independent of previous time periods, since the coefficients $c_{1,\tau}$ and $c_{2,\tau}$ depend on the initial state at the start of the interval, which in turn depends on the preceding interval.

B. Communication Attack Model

A broad class of attacks can be modeled mathematically by modifying the state-space equations. The change in the state-space modeled is based on the outcome of a stochastic process, as described in Section II-B. Two matrices $\mathbf{G}_\tau$ and $\mathbf{H}_\tau$, which contain random variables from the stochastic process, model the injection attack on measurements and commands respectively, while $\mathbf{M}_\tau$ models the defensive action. The modified state-space model is given by:

$$\mathbf{y}_\tau(t) = \mathbf{C}\mathbf{G}_\tau \mathbf{x}_\tau(t) \tag{15}$$

$$\mathbf{u}_\tau(t) = -\mathbf{H}_\tau \mathbf{K}\mathbf{C}\mathbf{G}_\tau \mathbf{x}_\tau(t) \tag{16}$$

$$\dot{\mathbf{x}}_\tau(t) = (\mathbf{A} - \mathbf{B}\boldsymbol{\Theta}_\tau \mathbf{H}_\tau \mathbf{K}\mathbf{C}\mathbf{G}_\tau)\mathbf{x}_\tau(t) \tag{17}$$

$$\mathbf{A}_\tau = \begin{pmatrix} -\frac{D_i}{M_i} & -\frac{1}{M_i} & \frac{1}{M_i} & 0 & 0 \\ 2\pi \sum_{j \neq i} T_{ij} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{T_{\mathrm{ch}_i}} & \frac{1}{T_{\mathrm{ch}_i}} & 0 \\ -\frac{1}{R_i T_{g_i}} - \frac{f(\theta_i) h_i g_i \beta_i K_{P_i}}{T_{g_i}} & -\frac{f(\theta_i) h_i K_{P_i}}{T_{g_i}} & 0 & -\frac{1}{T_{g_i}} & -\frac{f(\theta_i) h_i K_{I_i}}{T_{g_i}} \\ \beta_i & 1 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{5 \times 5} \tag{18}$$

Both $\mathbf{G}_\tau$ and $\mathbf{H}_\tau$ linearly scale the target variables. $\mathbf{G}_\tau$ modifies frequency deviation Δf measurements while $\mathbf{H}_\tau$ scales the load references setpoints for generators. In case of a positive detection, estimation accuracy for the expected command is given by $\mathbf{\Theta}_\tau$, which is a function of the reconstruction accuracy θ. $\mathbf{H}_i = [h_i]$ and $\mathbf{\Theta}_i = [f(\theta_i)]$ are scalars while $\mathbf{G}_i = \mathrm{diag}(g_i, 1, 1, 1, 1)$ is a diagonal matrix containing the scaling factors for the i-th BA area. Denoting $\mathbf{A}_\tau = \mathbf{A} - \mathbf{B}\mathbf{\Theta}_\tau \mathbf{H}_\tau \mathbf{K} \mathbf{C} \mathbf{G}_\tau$, the individual matrix elements in the modified system are shown below. The eigenvalues of $\mathbf{A}_\tau$ in (18) determine system behavior for various τ, depending on the outcome of the stochastic process.

REFERENCES

[1] T. S. Sidhu and Y. Yin, "Modelling and simulation for performance evaluation of IEC61850-based substation communication systems," *IEEE Trans. Power Del.*, vol. 22, no. 3, pp. 1482–1489, 2007.

[2] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams, and A. Hahn, "Guide to Industrial Control Systems (ICS) Security," National Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep., jun 2015. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf

[3] W. Wang and Z. Lu, "Cyber security in the Smart Grid: Survey and challenges," *Computer Networks*, vol. 57, no. 5, pp. 1344–1371, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1389128613000042

[4] C. C. Sun, A. Hahn, and C. C. Liu, "Cyber security of a power grid: State-of-the-art," *International Journal of Electrical Power and Energy Systems*, vol. 99, pp. 45–56, Jul. 2018.

[5] L. Jiang, W. Yao, Q. H. Wu, J. Y. Wen, and S. J. Cheng, "Delay-dependent stability for load frequency control with constant and time-varying delays," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 932–941, may 2012. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6080746

[6] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, "Preventing Time-Delay Switch Attack on Load Frequency Control in Distributed Power Systems," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 1176–1185, mar 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7352356/

[7] M. Khalaf, A. Youssef, and E. El-Saadany, "Joint Detection and Mitigation of False Data Injection Attacks in AGC Systems," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4985–4995, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8472173

[10] S. Cheung, B. Dutertre, M. Fong, U. Lindqvist, K. Skinner, and A. Valdes, "Using Model-based Intrusion Detection for SCADA Networks," in *Proceedings of the SCADA Security Scientific Symposium*. Miami Beach, FL: SRI International, jan 2007, pp. 1–12. [Online]. Available: http://www.csl.sri.com/papers/scadaIDS07/

[8] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, dec 2011. [Online]. Available: http://ieeexplore.ieee.org/document/6032057/

[9] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, pp. 1–33, may 2011. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1952982.1952995

[11] R. Mitchell and I. R. Chen, "Behavior-rule based intrusion detection systems for safety critical smart grid applications," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1254–1263, 2013.

[12] S. Pan, T. Morris, and U. Adhikari, "Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3104–3113, nov 2015.

[13] J. Hong and C. C. Liu, "Intelligent Electronic Devices with Collaborative Intrusion Detection Systems," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 271–281, jan 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8006250/

[14] Y. Zhang, L. Wang, W. Sun, R. C. Green, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 796–808, 2011.

[15] J. Hu, X. Yu, D. Qiu, and H. H. Chen, "A simple and efficient hidden Markov model scheme for host-based anomaly intrusion detection," *IEEE Netw.*, vol. 23, no. 1, pp. 42–47, 2009.

[16] X. Liu, M. Shahidehpour, Z. Li, X. Liu, Y. Cao, and Z. Li, "Power System Risk Assessment in Cyber Attacks Considering the Role of Protection Systems," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 572–580, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7439817/

[17] Y. W. Law, T. Alpcan, and M. Palaniswami, "Security games for risk minimization in automatic generation control," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 223–232, Jan 2015.

[18] L. Wei, A. I. Sarwat, and W. Saad, "Risk assessment of coordinated cyber-physical attacks against power grids: A stochastic game approach," in *2016 IEEE Industry Applications Society Annual Meeting*. Portland, OR, USA: IEEE, oct 2016, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/document/7731849/

[19] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, jun 2006.

[20] T. R. B. Kushal, "Decision analysis of cyber-physical resilience in power systems," Ph.D. dissertation, The Ohio State University, Columbus, OH, 2021.

[21] D. Wilson, J. Yu, N. Al-Ashwal, B. Heimisson, and V. Terzija, "Measuring effective area inertia to determine fast-acting frequency response requirements," *International Journal of Electrical Power and Energy Systems*, vol. 113, pp. 1–8, dec 2019.

[22] T. R. B. Kushal, K. Lai, and M. S. Illindala, "Risk-based Mitigation of Load Curtailment Cyber Attack Using Intelligent Agents in a Shipboard Power System," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4741–4750, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8449841