The final publication is available at Elsevier via http://dx.doi.org/https://doi.org/10.1016/j.ijepes.2019.05.014. © 2019. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

Mitigation of Cyber-Physical Attacks in Multi-Area Automatic Generation Control

Safoan Alhalali*, *Student Member, IEEE*, Christopher Nielsen†, *Member, IEEE*, Ramadan El–Shatshat†, *Senior Member, IEEE*

***Abstract*—In recent years, several cyber-attacks have been recorded against sensitive monitoring systems. Among them is the automatic generation control (AGC) system, a fundamental control system used in all power networks. AGC systems keep the network frequency at its desired value and maintain the tie-line power exchanges at their scheduled values. Motivated by the increasing need for robust and safe operation of AGCs, we introduce an attack resilient control scheme for the AGC system based on attack detection using state estimation. The proposed approach requires redundancy of sensors available at the transmission level in the power network and leverages recent results on attack detection using mixed integer linear programming (MILP). The proposed algorithm detects and identifies the sensors under attack in the presence of noise. The non-attacked sensors are then averaged and made available to the feedback controller. No assumptions about the nature of the attack signal are made. The proposed method is simulated using a large range of attack signals and uncertain sensors measurements.**

***Index Terms*—Power system network, automatic generation control, cyber security**

I. Introduction

In recent years, several cyber-attack incidents have been reported [1]. A detailed survey of different cyber-attack incidents was provided in [2]. Different strategies have been proposed in [3], [4] for a class of the actuator attacks. A detailed elaboration on cyber-attack incidents in power networks appears in [5], [6]. Little work has been conducted with respect to attack resilient measures that are used to detect, identify, and mitigate corrupted real-time measurements in the feedback loop of automatic generation control (AGC).

The accuracy and reliability of real-time measurements has a significant impact on system's real-time operation. In smart power grids, real-time measurements for AGC are transmitted using computer networks [7]. These computer networks might be an attractive space for cyber-attackers, e.g., disgruntled employees, insiders, nation states or terrorist organizations [8]. Through the computer network, the attacker can also learn the parameters of the system using the algorithm shown in [7]. The compromised measurements may lead to very severe and adverse effects on the management and control of a smart grid. For example, the corrupted measurements can cause a rapid decline in the system frequency that leads to trigger load shedding schemes or generators disconnecting. This is because the compromised measurements can mislead the AGC, which requires minimal supervision and intervention by human operators, to increase generation through a false impression of increasing demand. The imbalance between power generation and demand leads to deviation of the grid frequency from its nominal value. The channels through which the attacker can corrupt sensor measurements include the physical sensors, sensor data communication links, and data processing programs at the control center. Accessing and attacking geographically distributed physical sensors is tedious and hard to coordinate. However, hacking a computer program at the strongly protected control center is not impossible [1], [2].

A. Related Works

The work in [7] assumes that the frequency sensor is well protected and builds their detection model based on a comparison between the prediction of frequency and the measured one. Therefore, in case of losing the frequency measurement the detection algorithm will be no longer functioning, leaving great chance for the attacker to manipulate the system.

Many of the proposed detection and identification algorithms for security of AGCs [2], [8], [9], [10] use limited attack templates that cannot characterize real-world attackers well. In [8] the author proposed a statistical method based on maximizing the likelihood of detecting the attacked sensor. However, many of the statistical methods [11], [12], [13], including [8], have the problem of false positives and false negatives. The work of [8] has reported 5% false positives and false negatives. The proposed mitigation strategy in [7], [8] is, upon detecting an attack, to replace the sensor data with forecasting data. Such a strategy can cause large deviation from the nominal frequency due to forecasting error.

The work in [7] identifies the sensors under attack after 20 seconds from the onset of the attack. The mitigation algorithm proposed in [7] is based on neglecting the measurements from the sensors under attack. Therefore, the attacker can manipulate the AGC system during the 20 second period, causing at least some profit losses to the owner or even triggering the remedial action. The work in [14], [15] have formulated the state estimation problem using mixed integer linear programming (MILP). However, the authors assume that the system is noiseless. No mitigation solution has been proposed in [14], [15], [16], [17].

B. Contributions of this Work

In this paper we developed a new attack resilient scheme for single and multi-area AGC systems. The proposed approach

*Supported by the Libyan Scholarship Program.
†Supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors are with the Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo ON, N2L 3G1 Canada. {salhalal;cnielsen; raelshat}@uwaterloo.ca

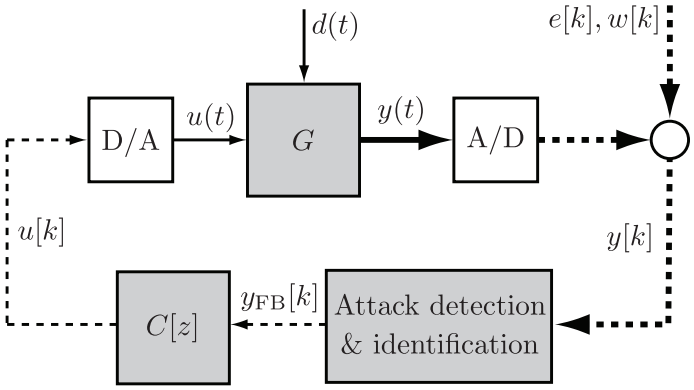

Fig. 1. Block diagram of a single AGC system (G) with the proposed attack detection and mitigation scheme. Dashed lines indicate discrete-time signals, solid lines represent continuous-time signals while thick lines indicate multi-sensor channels.

detects and identifies the sensors under attack on single area AGC systems leveraging the results presented in [15] and [17].

Once an attack has been identified, we employ a simple switching scheme to ensure that the AGC feedback loop continues to make its control decisions using uncompromised sensor data. We also characterize the degree of sensor redundancy needed in order to implement a robust solution and ensure uninterrupted service. We extend the above ideas to a multi-area AGC framework where the frequency and tie-line sensors may be attacked. We provide s-sparse observability analysis for AGC in single area and multi-area power networks. We provide numerical simulations demonstrating the effectiveness of the proposed approach in detecting and mitigating attacks as well as the feasibility of real-time implementation.

Figure 1 shows a block diagram of our proposed attack detection and mitigation scheme in the context of a single AGC system. The key idea is to have redundant sensors to measure the generator's (G's) frequency $y(t)$ so that, when an attack signal $e[k]$ corrupts the measurement, we correctly identify the un-attacked sensors and use this data in $y_{\mathrm{FB}}[k]$ to decide on the control signal $u[k]$. The signal $w[k]$ represents the measurement noise. The block $C[z]$ represent the controller while the blocks D/A and A/D are respectively represent the digital to analog converter and analog to digital converter.

We augment the conventional AGC system by adding more frequency sensors and model the attack vector to the measurement equation in Section II. The proposed detection and identification algorithm is presented in Section III. In Section IV we explain how the un-compromised sensors are fused in order to generate a reliable control signal. We extend these ideas to multi-area AGC systems in Section V while Section VI provides numerical simulation results.

II. Problem formulation for a single AGC

We present a standard continuous-time state-space representation of the linearised model of a single area AGC system. We use the conventional control architecture based on the linearised model. The state-space variable $x_1 := \Delta P_v$ equals the steam valve position command; the variable x_2 equals ΔP_m, the change in mechanical power; the state x_3 equals the frequency change $\Delta\omega$ of the power system network. The state vector of the AGC system is $x := (x_1, x_2, x_3) \in \mathbb{R}^3$. The exogenous signals are the control input $u(t)$ which physically represents the generation difference and the disturbance $d(t)$ which physically represents load change $\Delta P_L(t)$. With these definitions the linearised AGC dynamics can be expressed in state-space form as

$$\dot{x}(t) = Ax(t) + Bu(t) + Ed(t). \tag{1}$$

Following [18], the matrices in (1) are given by

$$A = \begin{bmatrix} -1/\tau_g & 0 & 0 \\ 1/\tau_T & -1/\tau_T & 0 \\ 0 & 1/2H & -D/2H \end{bmatrix}, \quad B = \begin{bmatrix} 1/\tau_g \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

and $E^\top = \begin{bmatrix} 0 & 0 & 1/2H \end{bmatrix}$. The meaning of the physical parameters appearing in these matrices can be found in [18]. The signal available for feedback is $x_3(t)$ which is also the variable that must be regulated. Therefore, the system output is taken to be

$$y(t) = Cx(t) + w(t) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} x(t) + w(t) \tag{3}$$

where $w \in \mathbb{R}$ represents measurement noise and is assumed to take values in a *known* compact subset of the output space $\mathbb{R}$. In other words, there exists a known constant $\delta_w > 0$ such that for all $t \geq 0$, $|w(t)| \leq \delta_w$.

In practice, the controller is often designed for the continuous-time dynamics (1), (3) and then descretized for implementation. Assuming an ideal sample and an ideal zero-order hold at a sampling period of $T > 0$ seconds, and further assuming that $d(t) = \Delta P_L(t)$, and $w(t)$ are constant over a sampling period, the AGC evolves in discrete time according to the dynamics,

$$\begin{aligned} x[k+1] &= A_d x[k] + B_d u[k] + E_d d[k] \\ y[k] &= C_d x[k] + w[k] \end{aligned} \tag{4}$$

where $x[k] := x(kT)$ is the value of the state at time kT, $k \in \mathbb{Z}$. Similarly $y[k] := y(kT)$, $d[k] := d(kT)$, $w[k] := w(kT)$. The matrices in (4) are given by $A_d = \mathrm{e}^{AT}$, $B_d = \int_0^T \mathrm{e}^{A\tau} B \, \mathrm{d}\tau$, $E_d = \int_0^T e^{A\tau} E \, \mathrm{d}\tau$ and $C_d = C$. We assume that the sampling is not pathological so that the pairs (A_d, B_d), (C_d, A_d) remain, respectively, controllable and observable. A commonly used control law is proportional-integral (PI) control [18] given, in discrete-time, by

$$\begin{aligned} x_c[k+1] &= x_c[k] + Ty[k] \\ u[k] &= K_P y[k] + K_I x_c[k] \end{aligned} \tag{5}$$

where $x_c[k]$ is the controller state at time kT and K_P, K_I are, respectively, the proportional and integral gains. During normal operation, thanks to the integral action in the controller, the single area AGC system is able to reject piecewise constant disturbances while keeping the system frequency at its nominal value.

A. Sensor redundancy and attack model

As mentioned in the introduction, in this paper we assume that the AGC is equipped with more than one frequency sensor. This redundancy will be used to detect cyber-physical attacks

and take appropriate control action. As such, we no longer assume that the measured output is 1-dimensional as in (4), but rather that there are p frequency sensors. Let $\mathbf{y} \in \mathbb{R}^p$ denote the information available for feedback using redundant sensors. As in the single output case (4), each sensor is assumed to be corrupted by measurement noise which, with mild abuse of notation, we denote as a vector $w \in \mathbb{R}^p$. We continue to assume that there is a known constant $\delta_w > 0$ such that[1] for all $k \in \mathbb{Z}$, $\|w[k]\|_\infty \leq \delta_w$. To model a cyber-physical attack, we let $e \in \mathbb{R}^p$ denote an *attack vector* so that $e[k]$ represents an attack on the sensors at time $t = kT$. This attack vector appears as an additive disturbance to the measured variables so that the AGC model, with redundant measurements and attacks, is given by

$$\begin{aligned} x[k+1] &= A_d x[k] + B_d u[k] + E_d d[k] \\ \mathbf{y}[k] &= \mathbf{C} x[k] + w[k] + e[k] \end{aligned} \tag{6}$$

where $\mathbf{C} \coloneqq (1_p \otimes C_d)$, $\otimes$ denotes the Kronecker product and $1_p \in \mathbb{R}^p$ is the vector of all ones. The first problem considered in this paper, under the assumption that we have sufficient redundancy in the sensors, is to identify which sensors have been corrupted by an attack and then use the non-attacked measurements as inputs to the control law (5) in order to ensure reliable operation of the AGC. In Section V we extend this idea to multi-area automatic generation control.

III. Attack detection

As stated in the introduction, we leverage the results in [15] and [17] to detect attacks on AGC systems. In this section, we summarize how the attack detection scheme works in the context of an AGC. We start by determining the maximum number of sensors that can be simultaneously attacked while still retaining the ability to detect the attacks when there is no noise and no load change. To do this we need some notation. If $\mathbb{N}_p \coloneqq \{1, \dots, p\}$ and $\mathcal{K} \subseteq \mathbb{N}_p$, then $\mathcal{K}^c \coloneqq \mathbb{N}_p \backslash \mathcal{K}$ denotes the complement of $\mathcal{K}$ in $\mathbb{N}_p$. Let $P_{\mathcal{K}} : \mathbb{R}^p \to \mathbb{R}^{|\mathcal{K}|}$ be the linear map which takes a vector $y \in \mathbb{R}^p$ and removes the rows in $\mathcal{K}^c$.

Definition III.1 ([19]). *A pair $(\mathbf{C}, A)$ is **s-sparse observable** if, for every set $\mathcal{K} \subset \mathbb{N}_p$ of cardinality s, the pair $(P_{\mathcal{K}^c}\mathbf{C}, A)$ is observable.*

We assume that the number of compromised sensors in (6) is less than or equal to $q_{\max}$, the maximum number of sensors for which the system's state can be recovered when there is no noise and no load changes.

Lemma III.2 ([15]). *The maximum number $q_{\max}$ of attacked sensors for which the state of (6) can be recovered when there is no noise and no disturbance equals the largest $s \in \mathbb{N}_p$ for which the pair $(\mathbf{C}, A_d)$ is $2s$-sparse observable.*

From this lemma we can immediately quantify the minimum degree of redundancy needed for frequency control of an AGC.

Proposition III.3. *The maximum number of sensors which can be attacked while retaining the ability to estimate the state of the single AGC system when there is no noise and no load changes equals $\lfloor (p-1)/2 \rfloor$.*

Proof. Observe that the pair (C, A_d) is observable. Therefore the pair $(P_{\mathcal{K}^c}\mathbf{C}, A_d)$ is observable as long as $P_{\mathcal{K}^c}$ is not equal to the zero map, i.e., $\mathcal{K}^c \neq \varnothing$. This in turn implies that $|\mathcal{K}| \leq p-1$. Setting $2q_{\max} = |\mathcal{K}|$ and solving for $q_{\max}$ gives the desired result. □

TABLE I
Number of frequency sensors p and maximum number of attacked sensors $q_{\max}$.

p	1	2	3	4	5	6	7	8	9	10	11
$q_{\max}$	0	0	1	1	2	2	3	3	4	4	5

Table I gives an interpretation of Proposition III.3. Given an expected maximal number of sensors $q_{\max}$ that can be attacked, we can choose the number of frequency sensors that the AGC should be equipped with.

Remark III.4. *The criteria for selecting the degree of sensor redundancy, from point of view of the operator, is based on a careful cost benefit analysis. Given the huge impact, both socially and economically, of an attack, having a large number of sensors may be justifiable. The full economic analysis of such decision is outside the scope of this paper and merits further research.*

Remark III.5. *A conventional AGC system can be attacked by a small team of attackers. The proposed control scheme introduces more obstacles to the attackers. By requiring that multiple sensors to be compromised simultaneously, the proposed scheme requires a larger, more well-funded and skilled team to overcome the defense.*

A. Identifying attacked sensors

The support of the attack vector e at time k is $\operatorname{supp}(e[k]) = \{i \in \mathbb{N}_p : e_i[k] \neq 0\}$, where e_i is the ith component of e. The symbol $\|e[k]\|_0 = |\operatorname{supp}(e[k])|$ denotes the number of non-zero entries in the vector $e[k]$. Let $\mathcal{K} \coloneqq \bigcup_{k \in \mathbb{Z}} \operatorname{supp}(e[k])$ and, in light of Proposition III.3, we assume throughout that[2] $|\mathcal{K}| \leq q_{\max} = \lfloor (p-1)/2 \rfloor$. To identify the sensors that have been attacked we collect the last N output measurements and solve an MILP at every time step. The value used for N is naturally taken to be equal to the dimension of the system's state space. Let $\tilde{\mathbf{y}}_i[k] \coloneqq \begin{bmatrix} P_i \tilde{\mathbf{y}}[k-N+1]^\top & \cdots & P_i \tilde{\mathbf{y}}[k]^\top \end{bmatrix}^\top \in \mathbb{R}^N$ be the vector which maintains the last N sensor measurements from sensor $i \in \mathbb{N}_p$, compensated for the inputs applied during this interval, i.e,

$$\begin{aligned} \tilde{\mathbf{y}}[k-N+1] &\coloneqq \mathbf{y}[k-N+1] \\ \tilde{\mathbf{y}}[n] &\coloneqq \mathbf{y}[n] - \sum_{j=0}^{n-k+N-2} \mathbf{C} A_d^j B_d u[n-j-1], \end{aligned}$$

where $n \in \{k-N+2, \dots, k\}$. Define, for output $i \in \mathbb{N}_p$, $O_i \coloneqq$

[1] If $x \in \mathbb{R}^n$, then $\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}$.

[2] This assumption can be weakened to state that, over any time interval of length N, the number of attacked sensors is less than $q_{\max}$.

$\begin{bmatrix}(P_i\mathbf{C})^\top & (P_i\mathbf{C}A)^\top & \cdots & (P_i\mathbf{C}A^{N-1})^\top\end{bmatrix}^\top \in \mathbb{R}^{N\times n}$ and let $\mathbf{e}_i[k] \coloneqq \begin{bmatrix}e_i[k-N+1]^\top & \cdots & e_i[k]^\top\end{bmatrix}^\top \in \mathbb{R}^N$. The vector $\mathbf{e}_i$ represents the attack values injected into sensor i over the last N time steps. Similarly, define $\mathbf{w}_i[k] \coloneqq \begin{bmatrix}w_i[k-N+1]^\top & \cdots & w_i[k]^\top\end{bmatrix}^\top \in \mathbb{R}^N$ as the list of measurement noise over the last N time steps at sensor i. Then we can express the last N measurements obtained from output $i \in \mathbb{N}_p$ as

$$\tilde{\mathbf{y}}_i[k] = O_i x[k-N] + \mathbf{e}_i[k] + \mathbf{w}_i[k]. \tag{7}$$

Following [20], in this expression we have absorbed the effect of the disturbance d into the measurement noise terms and therefore, if necessary, increasing δ_w so that $\|w[k]\|_\infty \leq \delta_w$ continues to hold. Since the open-loop matrix (2) is Hurwitz, this approach is not overly conservative. Next we stack each of vectors, corresponding to each output, in (7) to define $\mathbf{Y}[k] \coloneqq \begin{bmatrix}\tilde{\mathbf{y}}_1[k]^\top & \cdots \tilde{\mathbf{y}}_p[k]^\top\end{bmatrix}^\top$, $\mathbf{E}[k] \coloneqq \begin{bmatrix}\mathbf{e}_1[k]^\top & \cdots \mathbf{e}_p[k]^\top\end{bmatrix}^\top$, $\mathbf{W}[k] \coloneqq \begin{bmatrix}\mathbf{w}_1[k]^\top & \cdots \mathbf{w}_p[k]^\top\end{bmatrix}^\top$. Each of these are pN-dimensional real vectors. Finally, define $O \coloneqq \begin{bmatrix}O_1^\top & \cdots & O_p^\top\end{bmatrix}^\top \in \mathbb{R}^{pN\times n}$ so we can compactly write

$$\mathbf{Y}[k] = Ox[k-N] + \mathbf{E}[k] + \mathbf{W}[k]. \tag{8}$$

By assumption, the measurement noise is uniformly bounded $\|w[k]\|_\infty \leq \delta_w$ and so the vector $\mathbf{W}[k]$ is restricted to the compact 'box' $\Omega \coloneqq [-\delta_w, \delta_w]^{pN} \subset \mathbb{R}^{pN}$.

Following the notation from [21], define

$$\begin{aligned}\|\mathbf{E}[k]\|_{2,0} &\coloneqq \sum_{i=1}^{p} \mathbb{I}(\|\mathbf{e}_i[k]\|_2 > 0),\\ \|\mathbf{E}[k]\|_{2,1} &\coloneqq \sum_{i=1}^{p} \|\mathbf{e}_i[k]\|_2,\end{aligned} \tag{9}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. The number $\|\mathbf{E}[k]\|_{2,0}$ equals the number of sensors that have been attacked over the last N time steps while $\|\mathbf{E}[k]\|_{2,1}$ represents the cumulative size of the attacks. Consider the following optimization problem:

Problem 1.

$$\begin{aligned}\underset{\mathbf{E},x[k-N]}{\text{minimize:}}\ & \|\mathbf{E}[k]\|_{2,0}\\ \text{subject to:}\ & \mathbf{Y}[k] - Ox[k-N] + \mathbf{E}[k] = \mathbf{W}[k]\\ & \mathbf{W} \in \Omega.\end{aligned}$$

△

In the absence of noise, i.e., when $\Omega = 0 \in \mathbb{R}^{pN}$, if the number of attacked sensors $|\mathcal{K}|$ is less than or equal to $q_{\max} = \lfloor (p-1)/2 \rfloor$, then the values of the minimizing decision variables for Problem 1 are the state $x[k-N]$ and the actual attack vector $\mathbf{E}^*[k]$ [15]. In the presence of noise, the minimizing variables aren't necessarily $(x[k-N], \mathbf{E}^*[k])$.

Problem 1 involves combinatorial optimization and can be solved using MILP solvers. However, solving this problem is NP–hard in the general case which limits its use to smaller size systems. We will see that, since our AGC system with $n = 3$ state, $N = 3$ time steps and $p = 3$ frequency sensors, solving Problem 1 is feasible in real-time applications. Another approach is to convexify Problem 1 and solve the following instead:

Problem 2.

$$\begin{aligned}\underset{\mathbf{E},x[k-N]}{\text{minimize:}}\ & \|\mathbf{E}[k]\|_{2,1}\\ \text{subject to:}\ & \mathbf{Y}[k] - Ox[k-N] - \mathbf{E}[k] = \mathbf{W}[k]\\ & \mathbf{W} \in \Omega.\end{aligned}$$

△

This problem can be solved efficiently but is only effective at detecting relatively large attacks.

B. Performance

In [17], the authors provide a bound on the error between the true value of $x[k-N]$ and its estimated value for both of the aforementioned optimization problems. Let $(\tilde{x}_{2,0}, \tilde{\mathbf{E}}_{2,0})$ be the minimizing values for the decision variables for Problem 1 and let $(\tilde{x}_{2,1}, \tilde{\mathbf{E}}_{2,1})$ be the minimizing values for the decision variables for Problem 2. Let $\mathbf{E}^*$ denote the true attack vector over the last N time steps. Define the errors for Problem 1 $\Delta x_{2,0} \coloneqq \tilde{x}_{2,0} - x[k-N]$, $\Delta\mathbf{E}_{2,0} \coloneqq \tilde{\mathbf{E}}_{2,0} - \mathbf{E}^*$ as well as the errors $\Delta x_{2,1} \coloneqq \tilde{x}_{2,1} - x[k-N]$, $\Delta\mathbf{E}_{2,1} \coloneqq \tilde{\mathbf{E}}_{2,1} - \mathbf{E}^*$ the errors for Problem 2. By first computing an error bound on $\Delta x_{2,0}$ and $\Delta x_{2,1}$, one can prove the existence of constants D_j^i, $i \in \{0,1\}$, $j \in \mathbb{N}_p$ such that if

$$\mathbb{I}\big(\|\tilde{\mathbf{e}}_j[k]\| > D_j^i\big), \tag{10}$$

where $\tilde{\mathbf{e}}_j[k]$ is the jth block vector in either $\tilde{\mathbf{E}}_{2,0}$ (when $i = 0$) or $\tilde{\mathbf{E}}_{2,1}$ (when $i = 1$), then sensor j has been attacked over the last N time steps. In this paper we use the integer decision variable $\mathbf{E}$ in Problem 1 as our estimate of which sensors were attacked over the last N time steps. In other words, we will say that sensor j has been attacked over the last N times steps if

$$\mathbb{I}(\|\tilde{\mathbf{e}}_j[k]\|_2 > 0) \tag{11}$$

While there is no guarantee that the policy (11) will correctly detect the attacked sensor, simulations suggest that this approach is able to identify sensors under attack even when the attacks are small in magnitude. In the case where, due to computational efficiency considerations, one instead solves Problem 2, the policy (10) (using D_j^1) guarantees no false positive attack detections if the sufficient condition [17, Eqn. (29)]] is satisfied. Unfortunately, it can be shown using the parameters in Table II that the single AGC system does not satisfy the aforementioned sufficient condition. Therefore, the policy (10) is neither reliable nor effective for small magnitude of attack[3]. Since the bound computed in [17] is very conservative, due to several applications of the triangular inequality in its derivation, a huge range of small magnitude of attacks elude the proposed attack detector.

[3]Based on the parameters of the single AGC system shown in Table II $D_i^{\tilde{e}^{l_0}} = 1.2077$, and $D_i^{\tilde{e}^{l_1}} = 0.0049$.

TABLE II
AGC PARAMETERS IN SINGLE AREA SYSTEM [8]

D	R	K_I	$H(s)$	$\tau_g(s)$	$\tau_T(s)$
0.8	0.05	7	5	0.2	0.5

IV. ATTACK MITIGATION

At each time step k, the proposed attack mitigation strategy for a single AGC system with $p \geq 3$ frequency sensors is the as follows.

1) Fix $N \geq 3$ to be the number of steps over which we aim to detect attacks. The lower bound of 3 comes from the dimension of the AGC's state-space.
2) If $k \geq N-1$, solve the mixed integer linear program Problem 1. Let $\tilde{\mathbf{E}}_{2,0}[k]$ denote the value of the decision variable returned by the solver.
3) Set
$$s[k] \coloneqq 1_p - \begin{bmatrix}\mathbb{I}(\|\tilde{\mathbf{e}}_1[k]\|_2 > 0) & \dots & \mathbb{I}(\|\tilde{\mathbf{e}}_p[k]\|_2 > 0)\end{bmatrix}^\top$$
where $\tilde{\mathbf{e}}_j[k]$ is the jth block vector in $\tilde{\mathbf{E}}_{2,0}[k]$. The ith component of $s[k]$ is 1 if we have not detected an attack on sensor i over the last N time steps. Otherwise the ith component equals zero.
4) Take the average of the un-attacked sensor readings as the information available for feedback
$$y_{\text{FB}}[k] \coloneqq \frac{s^\top[k]}{s^\top[k]1_p}\mathbf{y}[k]. \tag{12}$$
5) Update the discretized PI control signal
$$\begin{aligned} x_c[k+1] &= x_c[k] + T y_{\text{FB}}[k] \\ u[k] &= K_P y_{\text{FB}}[k] + K_I x_c[k] \end{aligned} \tag{13}$$
and return to Step 2 at the next sample instant.

V. EXTENSION TO MULTI-AREA AGC SYSTEMS

Conventional multi-area AGC is based upon tie-line bias control where each area tends to reduce the area control error (ACE) to zero [18].

Consider a power network with n areas represented as a set of vertices $\mathbb{N}_n = \{1, \dots, n\}$ and overhead or underground lines (tie- lines) represented by a set of edges $\mathsf{E} \subseteq \mathsf{V} \times \mathsf{V}$. The neighbours of area i are defined as $\mathcal{N}_i \coloneqq \{j \in \mathbb{N}_n : (i,j) \in \mathsf{E}\}$. The neighbours of area i are simply the areas connected to it via tie-lines. If $(i,j) \in \mathsf{E}$, then ΔP_{ij} represents a deviation from the scheduled exchanges between areas i and j. The variable $\Delta\omega_i$ represents the deviation from the nominal frequency value for area i. With this notation, the area control error for area i consists of a linear combination of frequency and its neighbouring tie-line error $\text{ACE}_i \coloneqq \beta_i \Delta\omega_i + \sum_{j\in\mathcal{N}_i} \Delta P_{ij}$. The area bias β_i determines the amount of interaction during a disturbance in the neighbouring areas. To model the interconnection with its neighbours, we modify the single AGC continuous-time model (1) as follows. For simplicity assume that all the AGCs have the same physical constants. Let $r_i \coloneqq |\mathcal{N}_i|$ and $n_i \coloneqq 3 + r_i$. The three comes from the original state variables in (1) and the $|\mathcal{N}_i|$ extra states come from the interconnections. The model of AGC i in the multi-area AGC setup is then given by

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + E_i d_i(t) \tag{14}$$

with $A_i \in \mathbb{R}^{n_i \times n_i}$, $B_i \in \mathbb{R}^{n_i}$, $E_i \in \mathbb{R}^{n_i \times (1+r_i)}$. Specifically, letting $N_i \coloneqq e_3 1_{r_i}^\top \in \mathbb{R}^{3\times r_i}$, where $e_3 = \begin{bmatrix}0 & 0 & 1\end{bmatrix}^\top$ we have that

$$A_i = \begin{bmatrix} A & \frac{1}{2H}N_i \\ P_s N_i^\top & 0_{r_i \times r_i} \end{bmatrix}, B_i = \begin{bmatrix} B \\ 0_{r_i \times 1} \end{bmatrix}, E_i = \begin{bmatrix} E & 0_{3\times r_i} \\ 0_{1\times r_i} & -P_s I_{r_i} \end{bmatrix} \tag{15}$$

where A, B, E are the same matrices as in the single AGC system (1). The constant P_s is the synchronizing power coefficient, and equals the slope of the power angle curve at the initial operating angle. The first three state variables are the same as before while the last $r_i = |\mathcal{N}_i|$ state variables are ΔP_{ij}, i.e., the deviation from the scheduled exchanges between areas i and j where $j \in \mathcal{N}_i$. The control signal $u_i(t)$ is the same as in the single AGC model and the disturbance vector $d_i(t) = (\Delta P_L(t), \Delta\omega_{j_1}(t), \dots, \Delta\omega_{j_{r_i}}(t))$ is the load change on AGC i followed by the frequency change of each of the AGC's neighbours.

In the conventional setup, AGC i has access to its own frequency measurement $\Delta\omega_i$ as well as ΔP_{ij} for each $j \in \mathcal{N}_i$. Therefore the information available for feedback is given by

$$y_i(t) = C_i x_i(t) + w_i(t) \tag{16}$$

where

$$C_i = \begin{bmatrix} C & 0_{1\times r_i} \\ 0_{r_i \times 3} & I_{r_i} \end{bmatrix} \in \mathbb{R}^{(1+r_i)\times n}. \tag{17}$$

Once again the measurement noise $w_i(t)$ is assumed to be bounded in a known compact set. The conventional control scheme [18] is to use the PI controller $u_i(t) = K_{P,i} y_{i,1}(t) + K_{I,i}\int_0^t \begin{bmatrix}\beta_i & 1 & \cdots & 1\end{bmatrix} y_i(\tau)\mathrm{d}\tau$ where $y_{i,1}$ is the first component of y_i, i.e., AGC i's own frequency measurement. The scalers $K_{P,i}$, $K_{I,i}$ are, respectively, the proportional and integral gains of AGC i. Satisfactory performance is achieved by setting $\beta_i = K_{P,i} + D_i$ [18]

A. Proposed Attack-resilient State Estimation for the Multi Areas AGC System

Descretize the continuous time model (14), (16) of AGC i in the same manner as (6). The attack vector appears as an additive disturbance to the measured variables so that the AGC i model with redundant measurements and attacks is given by

$$\begin{aligned} x_i[k+1] &= A_{i_d} x_i[k] + B_{i_d} u_i[k] + E_{i_d} d_i[k] \\ \mathbf{y}_i[k] &= \mathbf{C_i} x_i[k] + w_i[k] + e_i[k] \end{aligned} \tag{18}$$

where $\mathbf{C_i}^\top \coloneqq \begin{bmatrix}1_{p_1} \otimes C_{i,1} & 1_{p_2} \otimes C_{i,2} & \cdot\cdot & 1_{p_{r_i+1}} \otimes C_{i,r_i+1}\end{bmatrix}$ The first p_1 rows in $\mathbf{C_i}$ represent the redundant frequency measurements, and $1_{p_{j+1}} \otimes C_{i,j+1}$, where $j \in \{1, \cdot\cdot, r_i\}$, represent jth tie-line redundant measurements. Once again the sampling is assumed to be non-pathological. We make the mild assumption that all interconnected AGCs are sampled at the same rate $T > 0$ and that all sample and hold operators

in the network are synchronised. The discretized PI controller for AGC i in state-space form is

$$\begin{aligned} x_{c,i}[k+1] &= x_{c,i}[k] + Ty_i[k] \\ u_i[k] &= K_{P,i}y_{i,1}[k] + K_{I,i}\begin{bmatrix}\beta_i & 1 & \cdots & 1\end{bmatrix} x_{c,i}[k]. \end{aligned} \tag{19}$$

The first issue considered in this section, under the assumption that we have sufficient redundancy in the sensors, is to identify which sensors have been corrupted by an attack and then use the non-attacked measurements as inputs to the control law (19). To understand how many attacks can be tolerated, we need to understand the s-sparse observability of (14), (16).

Proposition V.1. *The pair $(P_{\mathcal{K}}C_i, A_i)$, where A_i is given in (15), C_i is given by (17) and $\mathcal{K} \subseteq \mathbb{N}_{r_i+1}$, is observable if, and only if*

(i) $\mathcal{K} \neq \varnothing$, *and*
(ii) $|\mathcal{K} \cap \{2, \ldots, r_i+1\}| \geq r_i - 1$.

The conditions of this proposition can be intuitively interpreted as saying that AGC i is observable so long as it's only missing information from *at most* one of its neighbours.

Proof. We start by understanding the structure of the unobservable subspace of the pair (C_i, A_i). To simplify notation let c_j denote the jth row of C_i. Then $\operatorname{Ker} c_1 \subseteq \operatorname{Ker} c_j A_i$ for $j \geq 2$. This can be shown by noting that $c_j A_i = c_1$. It follows that $\operatorname{Ker}(c_1 A_i^{k-1}) \subseteq \operatorname{Ker}(c_j A_i^k)$ for $k \geq 1$, $2 \leq j \leq r_i + 1$. Therefore the unobservable subspace of the pair (C_i, A_i) can be written as follows,

$$\begin{aligned} \mathcal{N}_i &= \bigcap_{k=1}^{n_i} \operatorname{Ker}\left(C_i A_i^k\right) = \bigcap_{k=1}^{n_i} \left(\bigcap_{j=1}^{r_i+1} \operatorname{Ker}\left(c_j A_i^k\right) \right) \\ &= \operatorname{Ker} C_i \cap \operatorname{Ker}\left(c_1 A_i\right) \cap \cdots \cap \operatorname{Ker}\left(c_1 A_i^{n_i-1}\right) \end{aligned}$$

where we have used $\operatorname{Ker}(c_1 A_i^{k-1}) \subseteq \operatorname{Ker}(c_j A_i^k)$ for $k \geq 1$ and $2 \leq j \leq r_i+1$ to obtain the last equality. We can re-write the last line in the form $\mathcal{N}_i = \left(\bigcap_{j=2}^{r_i+1} \operatorname{Ker} c_j\right) \cap \mathcal{N}_{(c_1,A_i)}$, where $\mathcal{N}_{(c_1,A_i)}$ is the unobservable subspace of (c_1, A_i). We claim that the dimension of $\mathcal{N}_{(c_1,A_i)}$ equals $n_i - 4 = r_i - 1$. To show this, compute the first 5 rows of the observability matrix of (c_1, A_i)

$$\begin{bmatrix} c & 0 \\ cA & \gamma_1 1_{r_i}^\top \\ cA^2 + \gamma_2 c & \gamma_3 1_{r_i}^\top \\ cA^3 + \gamma_2 cA + \gamma_4 c & \gamma_5 1_{r_i}^\top \\ cA^4 + \gamma_2 cA^2 + \gamma_6 cA + \gamma_7 c & \gamma_8 1_{r_i}^\top. \end{bmatrix}$$

The various constants γ_i in this matrix are $\gamma_1 = 1/(2H)$, $\gamma_2 = r_i P_s \gamma_1$, $\gamma_3 = \gamma_1 cAc^\top$, $\gamma_4 = r_i\gamma_3$, $\gamma_5 = \gamma_1 c\left(A^2 + \gamma_2 I\right)c^\top$, $\gamma_6 = r_i P_s \gamma_3$, $\gamma_7 = r_i P_s \gamma_5$, $\gamma_8 = \gamma_1 c\left(A^3 + \gamma_2 A\gamma_2 I\right)c^\top$. Performing elementary row reduction on this matrix, then, using the Cayley-Hamilton theorem before doing further row reduction, we obtain

$$\begin{bmatrix} c & 0 \\ cA & \gamma_1 1_{r_i}^\top \\ cA^2 & cAc^\top/(2H) 1_{r_i}^\top \\ 0 & \gamma_9 1_{r_i}^\top \\ 0 & \gamma_{10} 1_{r_i}^\top. \end{bmatrix}$$

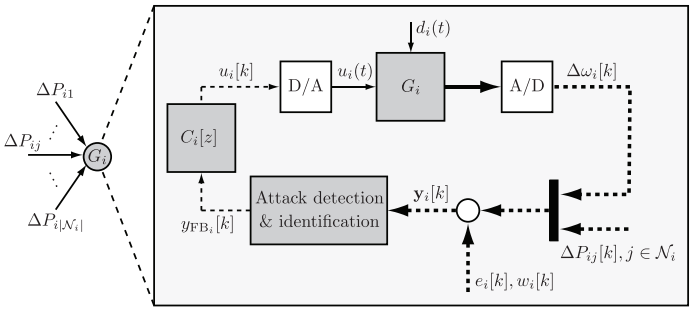

Fig. 2. Block diagram of AGC i's (G_i) attack detection and mitigation scheme. The attacker's signal is e_i while ΔP_{ij} is the deviation from the scheduled exchanges between areas i and j.

The coefficient γ_9 is guaranteed to be positive because $cA^2c^\top/(2H)$ is positive and, since A is Hurwitz, the coefficients of its characteristic polynomial are all positive. Since (c, A) is observable (cf. Section III), we conclude that this matrix has rank 4 and that the dimension of $\mathcal{N}_{(c_1,A_i)}$ equals $n_i - 4 = r_i - 1$ as claimed. We can therefore conclude that $\mathcal{N}_i = \{0\}$, i.e., (C_i, A_i) is observable, if and only if there are $r_i - 1$ tie-line measurements. This is precisely condition (ii) of the proposition statement while condition (i) ensures that the frequency sensor is measured in the case when the AGC only has one neighbour. □

Corollary V.2. *For $r_i = 1$, the maximum number of sensors which can be attacked when there is no noise and no disturbances while retaining the ability to estimate the state of the AGC i equals $\lfloor (p_1 + p_2 - 1)/2 \rfloor$.*

Proof. Based on Proposition V.1, for any sensor $j \in \mathbb{N}_{p_1+p_2}$ observe that the pair $(P_j C_i, A_{i_d})$ is observable. Therefore the pair $(P_{\mathcal{K}^c}\mathbf{C_i}, A)$ is observable as long as $P_{\mathcal{K}^c}$ is not equal to the zero map, i.e., $\mathcal{K}^c \neq \varnothing$. This in turn implies that $|\mathcal{K}| \leq p_1 + p_2 - 1$. Setting $2q_{\max} = |\mathcal{K}|$ and solving for $q_{\max}$ gives the desired result. □

Before, introducing Corollary V.3, for sake of clarity and without loss of generality, assume that the tie-line measurements p_{j+1} are ordered in increasing manner, where $p_2 \leq p_3 \leq \cdots \leq p_{r_i+1}$.

Corollary V.3. *For $r_i \geq 2$, the maximum number of sensors which can be attacked when there is no noise and no disturbances while retaining the ability to estimate the state of the AGC i equals $\lfloor (p_2 + p_3 - 1)/2 \rfloor$.*

Proof. Based on Proposition V.1, observe that the pair (C_i, A_i) is observable as long as the number of available tie-line measurements is at least equal to $r_i - 1$. Therefore the pair $(P_{\mathcal{K}^c}\mathbf{C_i}, A)$ is observable as long as the cardinality of $\mathcal{K}$ is less than $p_2 + p_3$. Setting $2q_{\max} = |\mathcal{K}|$ and solving for $q_{\max}$ gives the desired result. □

We now present the attack mitigation strategy for AGC in multi-area power network. To understand the strategy, note that each AGC in the network solves it's own version of Problem 1, *mutatis mutandis*, using its own model information. At each

time step k, the proposed attack mitigation strategy for AGC i is as follows.

1) Fix $N \geq 3+r_i$ to be the number of steps over which we aim to detect attacks. The lower bound of $3+r_i$ comes from the dimension of the AGC's state-space.
2) If $k \geq N-1$, solve the mixed integer linear program Problem 1. Let $\tilde{\mathbf{E}}_{2,0}[k]$ denote the value of the decision variable returned by the solver.
3) For the redundant measurements, we recursively define $\rho_0 := 0$, $\rho_j := \rho_{j-1} + p_j$ for $j \geq 1$ and then set

$$s_j[k] := 1_{p_j}^\top - \begin{bmatrix} \mathbb{I}\left(\|\tilde{\mathbf{e}}_{\rho_{j-1}+1}[k]\|_2 > 0\right) & \cdots & \mathbb{I}\left(\|\tilde{\mathbf{e}}_{\rho_j}[k]\|_2 > 0\right) \end{bmatrix}^\top. \tag{20}$$

where $\tilde{\mathbf{e}}_j[k]$ is the jth block vector in $\tilde{\mathbf{E}}_{2,0}[k]$. The ith component of $s_j[k]$ is 1 if we haven not detected an attack on sensor i over the last N time steps, where $j \in \{1, \cdot\cdot, r_i+1\}$. Otherwise the ith component equals zero.

4) Take the average of the un-attacked sensor readings as the information available for feedback

$$y_{\mathrm{FB}i}[k] := \mathbf{M} y_i[k]. \tag{21}$$

where, $\mathbf{M} := \mathrm{diag}\left(\frac{s_1^\top[k]}{s_1^\top[k]1_{p_1}}, \cdots, \frac{s_{r_i+1}^\top[k]}{s_{r_i+1}^\top[k]1_{p_{r_i+1}}}\right)$.

5) Update the discretized PI control signal

$$x_{c,i}[k+1] = x_{c,i}[k] + T y_{\mathrm{FB}i}[k] \tag{22}$$

$$u_i[k] = K_{P,i} y_{\mathrm{FB}i,1}[k] + K_{I,i}\begin{bmatrix}\beta_i & 1 & \cdots & 1\end{bmatrix} x_{c,i}[k] \tag{23}$$

where, $y_{\mathrm{FB}i,1}[k]$ is the first element in $y_{\mathrm{FB}i}[k]$ and return to Step 2 at the next sample instant.

Of course, for this strategy to be feasible for real-time control of the AGC i, one must be able to solve Problem 1 within the scheduling constraints of the embedded controller. The real-time constraint becomes too restrictive as the number of neighbours increases. However, we show that for the use of a fully connected, 3 AGC system, Problem 1 can be solved in real-time. Most existing interconnected systems [22], [23] consist of at most three tie lines. This is due to economic and technical constraints discussed in [24], [25]. Therefore, from a practical point of view, the proposed algorithm can be applied to most multi-AGC systems in existence.

VI. SIMULATION RESULTS

We present and illustrate the simulation results of the proposed algorithm compared with legacy AGC in the single and three-area system. The parameters used in simulation are given, respectively, in Tables II and III. The parameters are in per unit on a common 1000 MVA base. We implement the proposed algorithm in MATLAB and time the execution of each iteration on a 3.6 GHz i7-7700 CPU. We use CVX solver (MOSEK) for solving Problem 1. We begin with the AGC in the single area with sampling period $T = 0.1\ s$. The number of steps N over which we aim to detect the attack is equal to 3 for the single AGC case. For simulation purposes, we consider the following attack signal on sensor i

$$e_i(k) = \begin{cases} \mathbf{a}(k - k_r^1 + 1), & \text{for}, k_r^1 \leq k \leq k_r^2 \quad \text{(ramp)} \\ \mathbf{a}, & \text{for}, k_c^1 \leq k \leq k_c^2 \quad \text{(constant)} \\ \mathbf{a} f(k), & \text{for}, k_p^1 \leq k \leq k_p^2 \quad \text{(pulse)} \\ \mathbf{a} \sim \mathcal{N}(\mu, \sigma), & \text{for}, k_k^1 \leq k \leq k_k^2 \quad \text{(random)} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where, $f(k)$, is a pulse wave with a 50% duty cycle and period $6T$. The parameters μ, σ are, respectively, the mean and the standard deviation of the Gaussian distribution. The parameter $\mathbf{a} \in \mathbb{R}$ controls the "size" of the attack signal. Informally, when $\mathbf{a}$ is small we call the attacks "stealthy". By small, we mean that $|\mathbf{a}|$ is sufficiently small so as to ensure that the frequency does not deviate from the safety region and thereby trigger remedial action

Remark VI.1. *We emphasize that our proposed algorithm is agnostic to the penetration path of the attacker. However, as pointed out in the introduction, in practice the most attractive and easiest space for the attacker is the computer network.*

Remark VI.2. *Our proposed algorithm makes no assumptions about the particular form of the attack signal. The choices of the attack made above were done solely for the purpose of simulation and are partially motivated by the classes of the attack signals in [2], [8], [9], [10].*

The load change ΔP_L is generated by scaling the steady state load by zero-mean Gaussian random variable of standard deviation 0.05 per unit (p.u.). We assume that the number of sensors available is equal to 3 ($p = 3$), so based on Proposition III.3, $q_{\max}$ is equal to one. Fig. 3 shows the frequency deviation of the grid in Hz against time (kT) in seconds. Fig. 3 shows the frequency deviation of the grid with and without the proposed algorithm, and compared with the original frequency deviation. Several kinds of attacks are applied during the first and the second shaded area to simulate the unpredictable behaviour of the attacker. For the stealthy attacks during the first shaded area, ramp, pulse, constant, and random signal have been initiated at $k_r^1 = 40$, $k_p^1 = 150$, $k_c^1 = 200$, and $k_k^1 = 250$ respectively. The ramp attack with $\mathbf{a} = -0.005$ lasts for 8.6 seconds ($k_r^2 = 125$). The pulse attack with $\mathbf{a} = -0.2$ lasts for 3.6 seconds ($k_p^2 = 185$). The constant with $\mathbf{a} = -0.2$, and random with $\mu = -0.2$, $\sigma = 0.1$, attack lasts for 2.6 seconds ($k_c^2 = 235$, and $k_k^2 = 285$). When the attacker is applying stealthy attacks (first shaded area), the frequency goes slightly above the true grid frequency value. This means that the generation is unnecessarily above the demand. Consequently, the grid is operated uneconomily, causing some profit loss to the owner. In the second shaded area, the ramp attack with $\mathbf{a} = 0.05$ is launched again at 30 seconds ($k_r^1 = 300$). The aim of this attack is to cause high deviation from the nominal frequency and trigger the remedial actions. The thresholds ϵ_L and ϵ_U shown in Fig. 3 are set to those for triggering remedial actions. We use $\epsilon_L = -0.5$ Hz and $\epsilon_U = 0.5$ Hz [7]. As shown in Fig. 3, the attacker manipulates the AGC system to make the grid frequency

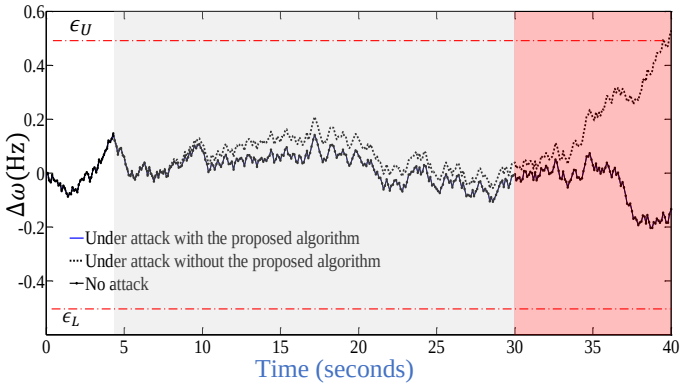

Fig. 3. The grid frequency deviation where the system is under stealthy attack (first shaded region) and under severe attack (second shaded region)

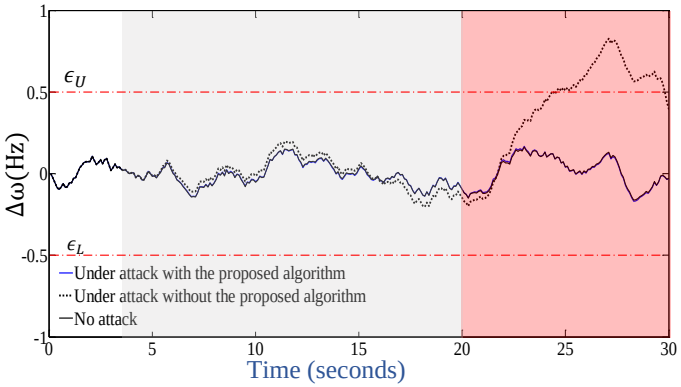

Fig. 4. Area 1 frequency deviation where the system is under stealthy attack (first shaded region) and under severe attack (second shaded region)

TABLE III
SYSTEM PARAMETERS [8]

Area 'i'	D	R	β	K_i	$H(s)$	$\tau_g(s)$	$\tau_T(s)$
Area 1	0.8	0.05	20.8	0.5	5	0.2	0.5
Area 2	0.9	0.0625	16.9	0.5	4	0.3	0.6
Area 3	0.9	0.0625	16.9	0.5	4	0.3	0.6

leave the safety region at time $37.5\,s$, while the AGC system equipped with proposed algorithm follows the not attacked frequency deviation. The proposed algorithm is able to detect and identify the sensors under attack in $0.07\,s$ (i.e., Problem 1 is solved using MOSEK in $0.07\ s$). The proposed algorithm is able to detect the attack of size $0.02\,\text{Hz}$ (i.e., $\mathbf{a} = 0.02$) once the attack is launched. Smaller sizes can be detected after 0.3 to 0.5 second. The proposed algorithm is able to detect and identify the sensors under attack even in the case of sudden load change or a high load fluctuation i.e., the proposed algorithm is able to detect and identify the sensors under attack when the load change ΔP_L is generated by scaling the highly fluctuated load by zero-mean Gaussian random variable with a standard deviation of 0.2 p.u.

Next, we test the AGC equipped with the proposed algorithm in a three-area configuration under various types of coordinated attacks with sampling period $T = 0.1\ s$. The number of steps N over which we aim to detect the attack is equal to 5 for the three-AGC case. For each AGC i, we assume that the number of frequency sensors available is equal to 3 ($p_1 = 3$), the number of first tie-line sensors available is equal to 3 ($p_2 = 3$), and the number of second tie-line sensors available is equal to 4 ($p_3 = 4$). Based on Corollary V.3, the maximum number of sensor that can be attacked is equal to 3 ($q_{\max} = 3$), however, simulations suggest that ($q_{\max} = 2$). This is due to the presence of noise and load disturbance. The proposed algorithm is able to detect and identify the tie-line sensors under attack even in the case of sudden load change or a high load fluctuation, while frequency sensors under attack can only be detected during steady-state load conditions.

In our simulation, the attacker misleads AGC 1 by decreasing the flow measurement from Area 1 to Area 2 by 0.03 p.u., and Area 1 to Area 3 by 0.03 p.u.. At the same time, the attacker keeps normal measurements according to the scheduled values to AGC in Areas 2 and 3 in order to prevent corrective action.

For the stealthy attacks during the first shaded area, pulse, constant, and random signal have been initiated at $k_p^1 = 40$, $k_c^1 = 80$, and $k_k^1 = 150$ respectively. The pulse attack with $\mathbf{a} = -0.03$ lasts for 3.6 seconds ($k_p^2 = 75$). The constant with $\mathbf{a} = -0.03$, last for 6.6 seconds ($k_c^2 = 145$). The random attacks with $\mu = -0.03$, and $\sigma = 0.01$ last for 2.6 seconds ($k_k^2 = 175$). In the second shaded area, the ramp attack with $\mathbf{a} = -0.01$ is launched at $20\,s$. As shown in Fig. 4, the attack will cause a remedial action as the frequency goes beyond the safe region at $24\,s$, while the AGC system equipped with proposed algorithm follows the not attacked frequency deviation. The proposed algorithm is able to detect and identify the sensors under attack in $0.09\ s$ (i.e., Problem 1 is solved in $0.09\ s$). The proposed algorithm is able to detect the attack on tie-line sensors of size 0.01 p.u. (i.e., $\mathbf{a} = 0.01$) once the attack launched , while for frequency sensors the size is $0.1\,\text{Hz}$ (i.e., $\mathbf{a} = 0.1$). Smaller sizes can be detected after 0.3 to 0.5 second.

The proposed algorithm was also tested on a system with twenty states (i.e., eighteen-area AGC system), and is able to detect and identify the sensors under attack in $0.51\ s$ (i.e., Problem 1 is solved using MOSEK in $0.51\,s$). However, since most existing multi-area AGC systems are based on three or two areas, we prefer to demonstrate the above explained example. We believe that the supercomputer of the power system operator (PSO) can solve Problem 1 in much faster time.

Remark VI.3. *Simulations, for both the single and mulit-AGC configurations, were conducted 10,000 times in order to demonstrate the reliability of the proposed algorithm. The simulations presented are representative of a typical load profile.*

VII. Conclusion

This paper developed an efficient algorithm for detecting, identifying and mitigating cyber-physical attacks for single and multi-area AGC systems. The proposed algorithm leverages a MILP-based state estimation procedure introduced in [15], and adapted for systems with noise in [17]. We provide s-sparse observability analysis for AGC in single area and multi-area

power networks. We derived a key formula to compute the number of sensors needed versus the number of attacks that can be tolerated. We propose a mitigation procedure based on simple switching algorithm. Our analysis and algorithms are validated by simulations for AGC in single and three-area power network. The proposed algorithm is capable of providing an accurate detection of attacks and identification of the sensors under attack even in the case of sudden load change or high load fluctuation.

REFERENCES

[1] S. Karnouskos, "Stuxnet worm impact on industrial cyber-physical system security," in *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, Nov 2011, pp. 4490–4494. 1

[2] S. Sridhar and G. Manimaran, "Data integrity attacks and their impacts on scada control system," in *IEEE PES General Meeting*, July 2010, pp. 1–6. 1, 7

[3] X. Huang and J. Dong, "Reliable control policy of cyber-physical systems against a class of frequency-constrained sensor and actuator attacks," *IEEE Transactions on Cybernetics*, 2018. 1

[4] X. Huang, D. Zhai, and J. Dong, "Adaptive integral sliding-mode control strategy of data-driven cyber-physical systems against a class of actuator attacks," *IET Control Theory & Applications*, vol. 12, no. 10, pp. 1440–1447, 2018. 1

[5] Y. Xiang, L. Wang, and N. Liu, "Coordinated attacks on electric power systems in a cyber-physical environment," *Electric Power Systems Research*, pp. 156 – 168. 1

[6] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu, "Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 847–855, June 2013. 1

[7] R. Tan, H. H. Nguyen, E. Y. S. Foo, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, "Modeling and mitigating impact of false data injection attacks on automatic generation control," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1609–1624, July 2017. 1, 7

[8] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, March 2014. 1, 5, 7, 8

[9] A. Ashok, P. Wang, M. Brown, and M. Govindarasu, "Experimental evaluation of cyber attacks on automatic generation control using a cps security testbed," in *2015 IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5. 1, 7

[10] S. Bhowmik, K. Tomsovic, and A. Bose, "Communication models for third party load frequency control," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 543–548, Feb 2004. 1, 7

[11] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *2013 American Control Conference*, June 2013, pp. 3344–3349. 1

[12] Y. Mo, T. H. J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, Jan 2012. 1

[13] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems*, vol. 35, no. 1, pp. 93–109, Feb 2015. 1

[14] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, Nov 2013. 1

[15] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, June 2014. 1, 2, 3, 4, 8

[16] P. M. Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, "Cyber attack in a two-area power system: Impact identification using reachability," in *Proceedings of the 2010 American Control Conference*, June 2010, pp. 962–967. 1

[17] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, March 2017. 1, 2, 3, 4, 8

[18] H. Saadat, *Power system analysis*. McGraw-Hill, 1999. 2, 5

[19] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, Aug 2016. 3

[20] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *ACM/IEEE 5th International Conference on Cyber-Physical Systems*, Berlin, Germany, 2014, pp. 163–174. 4

[21] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, June 2010. 4

[22] S. Hoff. U.S. electric system is made up of interconnections and balancing authorities. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=27152 7

[23] Wikipedia. High-voltage direct current. [Online]. Available: https://en.wikipedia.org/wiki/High-voltage_direct_current/media/File:HVDC_Europe.svg 7

[24] "Assessing the value of power interconnections under climate and natural gas price risks," *Energy*, vol. 82, pp. 128 – 137, 2015. 7

[25] I. E. S. O. for the Deputy Minister of Energy. Ontario-quebec interconnection capability a technical review. [Online]. Available: http://www.ieso.ca/-/media/files/ieso/document-library/power-data/supply/intertiereport-20170508.pdf?la=en. 7