A Reinforcement Learning Approach to Undetectable Attacks Against Automatic Generation Control

Ezzeldin Shereen, *Member, IEEE*, Kiarash Kazari, *Student Member, IEEE*, and György Dán, *Senior Member, IEEE*

***Abstract*—Automatic generation control (AGC) is an essential functionality for ensuring the stability of power systems, and its secure operation is thus of utmost importance to power system operators. In this paper, we investigate the vulnerability of AGC to false data injection attacks that could remain undetected by traditional detection methods based on the area control error (ACE) and the recently proposed unknown input observer (UIO). We formulate the problem of computing undetectable attacks as a multi-objective partially observable Markov decision process. We propose a flexible reward function that allows to explore the trade-off between attack impact and detectability, and use the proximal policy optimization (PPO) algorithm for learning efficient attack policies. Through extensive simulations of a 3-area power system, we show that the proposed attacks can drive the frequency beyond critical limits, while remaining undetectable by state-of-the-art algorithms employed for fault and attack detection in AGC. Our results also show that detectors trained using supervised and unsupervised machine learning can both significantly outperform existing detectors.**

***Index Terms*—Automatic generation control, reinforcement learning, false data injection attack, power system security, unknown input observer, partially observable Markov decision process.**

I. Introduction

Maintaining the balance between electric power generation and demand is one of the main objectives in power system operation. An imbalance between generation and demand can cause the grid frequency to deviate significantly, which can cause physical damage to generators, trigger remedial actions such as load shedding, or even lead to nation-wide blackouts. To mitigate these effects, automatic generation control (AGC) [1], [2] is a control loop that is used by operators to set the generation output of generators in a power system, based on power and frequency measurements taken across the interconnected power system. In AGC, the power system is typically divided into several areas, and a separate AGC controller is deployed for each area. The AGC controller attempts to minimize the deviation of the measured power flows across certain transmission lines and the grid frequency from their expected values. This is typically achieved by minimizing the area control error (ACE) metric, which is a weighted sum of the two aforementioned quantities.

The operation of AGC is dependent on the accuracy and integrity of the deployed sensor measurements. Nevertheless, since modern power systems usually utilize insecure public communication networks, the AGC control loop is vulnerable to a wide range of cyber-attacks. One of the most studied attacks is the false data injection attack (FDIA) [3], in which the attacker uses the communication network to inject false measurements and transmit them to the control center, where the AGC controller typically resides. The false measurements could cause the AGC controller to issue incorrect dispatch commands to the generators, potentially leading to catastrophic consequences in the power system. Therefore, extensive surveillance of the AGC control loop (including the sensor measurements) is an important aspect of the security of any power system.

Conventional solutions for detecting FDIAs against AGC systems depend on simply monitoring the ACE value at each area [1]. However, these methods do not utilize information from the AGC system model. Therefore, a recent promising approach for FDIA detection is utilizing the unknown input observer (UIO) [4], which can accurately estimate the unknown system states affecting AGC operation given (1) the observed sensor measurements and (2) accurate knowledge of the power system topology and parameters. An attack or a fault in AGC operation will usually lead to high estimation residuals, which causes an alarm to be raised. This approach has shown great potential in detecting naively computed FDIAs, such as the scaling, ramp, and random attacks. Nevertheless, the vulnerability of UIOs to targeted FDIAs has not yet been fully explored.

In this paper we investigate the vulnerability of state-of-the-art FDIA detection methods in AGC systems using the framework of reinforcement learning (RL). The contributions of this paper are as follows.

Manuscript received 1 December 2022; revised 15 March 2023; accepted 18 June 2023. Date of publication 22 June 2023; date of current version 26 December 2023. This work was supported in part by the Swedish Civil Contingencies Agency (MSB) through the CERCES2 Project; in part by the Swedish Research Council under Project 2020-03860; and in part by the Swedish Foundation for Strategic Research through the CLAS Project under Grant RIT17-0046. Paper no. TSG-01798-2022. *(Corresponding author: Ezzeldin Shereen.)*

The authors are with the Division of Network and Systems Engineering, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: eshereen@kth.se; kkazari@kth.se; gyuri@kth.se).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2023.3288676.

Digital Object Identifier 10.1109/TSG.2023.3288676

1) We model the problem of finding stealthy FDIAs against AGC from the perspective of the attacker as a multi-objective partially observable Markov decision process (MO-POMDP) [5].
2) We develop a flexible reward function that allows the RL-based attack to maximize the attack impact, while keeping the detection metrics low.
3) We use extensive simulations to evaluate the proposed RL-based attacks and showcase their superiority over several baseline attacks in terms of attack impact and in terms of undetectability.

To the best of our knowledge, this is the first work that considers computing FDIAs that bypass state-of-the-art AGC attack detectors such as the UIO, and the first work that uses RL to compute such attacks against AGC.

The rest of this paper is organized as follows. Section II discusses previous work on attacks against AGC as well as their countermeasures. Section III presents our model of the AGC system, and the capabilities of the attacker. The problem of computing FDIAs is formulated as a MO-POMDP in Section IV. Section V evaluates the performance of the proposed attacks in terms of stability, impact detectability, as well as sensitivity to model inaccuracy. Finally, Section VI concludes the paper.

II. Related Work

Several recent works have investigated the security of AGC and its vulnerability to attacks. One of these attacks is the time delay attack (TDA), which delays the transmission of measurements sent from the sensors to the control center, or the control commands from the control center to the generators. Recent works have shown that TDAs can degrade the performance of AGC or even disable it [6], [7]. Nonetheless, the most studied attack is by far the false data injection attack (FDIA) where the attacker can compromise the measurements (e.g., power flows or frequency measurements), thus leading the AGC controller to send incorrect dispatch commands to the generators [8]. Such FDIAs have been shown to pose a severe threat to the system frequency [9].

A different line of work has developed improvements to FDIAs against AGC systems. A "mild" version of FDIA that gradually changes the measurements was proposed in [10], and it was shown that these attacks could still cause significant deviations in the system frequency. Authors in [11] developed an FDIA that maximizes the system frequency deviation, while keeping measurement perturbations within limits. Authors in [3] proposed an FDIA against AGC based on a model of the AGC system that minimizes the time until initiating remedial actions by the system operator. Notably, the proposed attack is able to bypass state-of-the-art bad data detection (BDD) methods used in power system state estimation. Another attack that can bypass BDD methods was proposed in [12]. In the first phase of the FDIA, the false measurements are designed to look like un-attacked cases, while the second phase finally drives the frequency beyond the safe range. More recently, [13] designed an attack that minimizes both the attack magnitude and the time until frequency violation, while keeping the attacked measurements and the ACE values within normal limits.

In response to the rising threat of FDIAs against AGC, their detection and mitigation have recently attracted significant research interest. The traditional approach is to monitor the ACE of each area [1], since an increase in the ACE could be a strong indicator of a system fault or an attack. Building on this simple intuition, other approaches utilized the ACE signals for attack detection in more complicated ways. Most notably, [9] proposed an anomaly detector that monitors the ACE values and compares them with predicted values based on load forecasts. Similarly, [14] used load forecasts to predict a range of normal ACE values, which can be used to both detect and compensate for FDIAs. Moreover, [15] used pattern recognition and supervised classification to predict whether the ACE signal is normal or attacked. Besides, [16] proposed two methods based on long short-term memory (LSTM) and discrete Fourier transform (DFT) to detect abnormalities in ACE time series. A multilayer perceptron (MLP) combined with feature selection was trained in [17] to distinguish between attacked and non-attacked ACE signals. Recently, [18] proposed a combination of fuzzy logic and neural networks for the detection of FDIAs, where the input data consisted of the ACE values as well as other measurements.

In contrast to the above ACE-dependent approaches, another common approach is the use of a mathematical model of AGC to detect FDIAs. The most commonly used models are the unknown input observers (UIOs) [19], [20]. In these works, a mathematical model of AGC is formulated and is used to perform a delayed estimation of the system states by observing the sensor measurements. By comparing the received measurements with the measurements expected based on the estimated state, faults and attacks against AGC could be detected. Developing on the basic idea of the UIO, [21] includes the attack as a part of the UIO model (i.e., as an unknown input) so that the model learns to estimate the system state as well as the attack, which allows for correcting corrupted measurements. Similarly, [22] designed a UIO for FDIA detection and combined it with a robust adaptive observer and the H_∞ technique to estimate and correct the attacks. A similar idea was developed in [23] for detecting attacks in a decentralized manner by building smaller models that utilize only state variables from a single area.

Several works considered other model-based approaches for detecting FDIAs in AGC systems. The approach in [24] combined state and attack estimation with attack compensation using observer-based output feedback control design. Authors in [25] considered the slightly different AGC problem in hybrid AC/HVDC grids, and designed a residual generator based on the system model to detect and recover attacks. A recent approach is proposed in [26], where the authors designed a set of sliding mode observers (SMOs) and Luenberger observers to detect FDIAs and identify the location of the attacks. Another model-based approach is investigated in [27], where the Kalman filter is proposed for FDIA detection in AGC systems. Moreover, [28] used the Kalman filter to estimate and correct the effect of the attack. Finally, contrary to most works which consider a linearized AGC system model, [29] took system non-linearities into consideration and proposed using a particle filter to detect FDIAs.

Other approaches for detecting FDIAs in AGC systems include [30], which applies dynamic watermarking to measurements fed to the AGC system to detect attacks. More recently, an ensemble method based on supervised machine learning applied to area-level features has been proposed for detecting FDIAs in a decentralized manner [31]. Similarly, authors in [32] proposed detecting FDIAs by training an unsupervised generative adversarial network (GAN) using historical measurement and load data. Another unsupervised technique is presented in [33], where FDIAs are detected using an autoencoder neural network with LSTM structured neurons.

Apart from the FDIA detection problem, many works focused on the problem of fault-tolerant control in AGC systems. Authors in [34] proposed FDIA-resilient control in AGC systems combining a Luenberger observer, an artificial neural network (ANN), and an extended Kalman filter. Moreover, [35] proposed an H_∞ controller for event-triggered AGC to control the system frequency under DoS attacks and FDIAs. Besides, an LSTM-based regression model was developed in [36] to predict and compensate for the FDIA signals in AGC. Finally, several research works used a game-theoretic approach to model the interaction between the system operator and the attacker. In the game formulated in [37], the attacker chooses between manipulating either half or all of the samples, and the operator chooses between two different configurations of a FDIA detector. In the game proposed in [38], the attacker could either attack both power and frequency measurements or only the frequency, while the defender could switch between two different FDIA detectors, namely support vector machine (SVM) and k-nearest neighbours (KNN).

A significant limitation of most of the aforementioned works studying AGC security is that they considered weak attack models. Simple FDIAs such as *ramp*, *pulse*, *step*, *scaling*, *sine*, *random*, and *replay* attacks [9], [10], [14], [17], [19], [23], [27], [28], [31] have been commonly utilized either (1) to quantify the impact of FDIAs on AGC systems, or (2) to evaluate FDIA detection approaches. However, several works proposed attacks that included a notion of stealthiness. FDIAs constructed in [11], [13], [25], [33] satisfied simple constraints, e.g., upper and lower bound constraints on the attacked measurements. The above naive attacks do not exploit any knowledge about the attacked AGC system model, nor about the detectors deployed by the system operator. Therefore, available detection methods in the literature could very effectively detect these naive FDIAs. However, it is unclear whether state-of-the-art detection methods could detect more intelligent attacks that can leverage insider information about the AGC system and the employed detectors. Therefore, a thorough study of the security of AGC w.r.t. to a strong attack model is highly needed.

It is also worth to note that very few works [3], [12] proposed attacks that can bypass bad data detection (BDD) techniques typically employed with power system state estimation. Nevertheless, these attacks are agnostic of any AGC-specific FDIA detectors, and should thus be detectable by those. Besides, although [32] considered attacks that are stealthy w.r.t. the AGC system model, computing those attacks requires access to the unknown inputs (e.g., loads) and the authors do not provide a clear FDIA computation procedure.

Going beyond the above works, constructing intelligent FDIAs against AGC systems could be regarded as an optimal sequential decision making problem, with the objective of maximizing the attack impact and stealthiness. To this end, we utilize the framework of reinforcement learning (RL) to compute FDIAs because (1) computing optimal attacks using traditional mathematical optimization tools could be infeasible for large and highly dynamic AGC systems, and (2) the RL approach only requires the availability of a system model and historical data, and the attack procedure could in principle be applied against other cyber-physical control systems.

Moreover, RL has been extensively used in various power systems optimization tasks. Several works have proposed AGC controllers using RL or multi-agent RL (MARL) instead of the widely used PI-controller [1], [2]. One of the first RL-based AGC controllers was proposed in [39], where the authors used the Q-learning algorithm [40] based on discretized actions (generation set points) and observations of either (1) the ACE values, or (2) the power-flow and frequency measurements. More recently, [41] treated AGC as a decentralized multi-agent problem (i.e., each area controller is considered as one agent) and utilized state and action discretization to use the double deep Q-network (DDQN) [42] algorithm with action discovery. MARL has also been used to solve the problem of automatic voltage control (AVC) [43], using the multi-agent deep deterministic policy gradient (MADDPG) [44] algorithm, which leverages centralized training and decentralized execution, and is able to deal with continuous actions and observations. Similarly, a multi-agent actor critic RL algorithm was proposed in [45] to solve the problem of voltage and frequency control in inverter-based microgrids. Finally, Q-learning has been proposed to compute FDIAs against power system state estimation [46]. Nevertheless, to the best of our knowledge, our work is the first work to consider RL-based attacks against AGC, including the question of detectability using state-of-the-art detectors.

III. System Model

A. Automatic Generation Control

We consider an interconnected power system consisting of N areas, connected by power transmission lines called tie lines. We denote by $P^{sch}_{i,j}$ the scheduled (planned) power flow from area i to area j across their corresponding tie line(s), by $P^{tie}_{i,j}$ the actual power flow from area i to area j, and by $\Delta P^{tie}_{i,j} = P^{tie}_{i,j} - P^{sch}_{i,j}$ the deviation from the scheduled values. We denote by f_i the AC frequency of area i, and its deviation from the nominal grid frequency (e.g., $f_0 = 60$ Hz) by $\Delta\omega_i = \frac{f_i - f_0}{f_0}$. Each area has one or more electric power generators whose generation levels are controlled by the AGC in order to keep the deviations of both the frequency and the tie line power flows close to zero, despite changes ΔP^L in the electrical loads in each area.

At time instant t, the evolution of the frequency deviation $\Delta\omega_i$ is given by the differential equation

$$\Delta\dot{\omega}_i(t) = \frac{1}{2H_i}\big(\Delta P_i^m(t) - \Delta P_i^{tie}(t) - \Delta P_i^L(t) - D_i\Delta\omega_i(t)\big), \tag{1}$$

where H_i is the inertia constant of generator i, ΔP_i^m is the deviation in the mechanical power output of generator i, $\Delta P_i^{tie}(t) = \sum_{j=1}^N \Delta P_{i,j}^{tie}(t)$, and D_i is the damping coefficient of generator i. The power flow on a tie line can be approximated by

$$\Delta\dot{P}_{i,j}^{tie}(t) = P_{ij}^s\big(\Delta\omega_i(t) - \Delta\omega_j(t)\big), \tag{2}$$

where P_{ij}^s is the synchronizing power coefficient between areas i and j [2].

To drive the power and frequency deviations back to zero, each area's generator governor adjusts the position of the turbine's steam valve P_i^v based on the differential equation

$$\Delta\dot{P}_i^v(t) = -\frac{1}{\tau_i^g}\left(\frac{1}{R_i}\Delta\omega_i(t) + \Delta P_i^v(t) - \Delta P_i^{ref}(t)\right), \tag{3}$$

where τ_i^g is the time constant of the governor in area i, R_i is the speed regulation (droop) coefficient of the generator, and ΔP_i^{ref} is the input reference power generation of area i supplied by AGC. Changing ΔP^v will in turn control the output mechanical power ΔP^m as

$$\Delta\dot{P}_i^m(t) = -\frac{1}{\tau_i^t}\big(\Delta P_i^m(t) - \Delta P_i^v(t)\big), \tag{4}$$

where τ_i^t is the turbine time constant of area i.

To regulate the frequency and the tie line power flows, the AGC controller is typically implemented as a PI-controller that controls ΔP^{ref} using

$$\Delta\dot{P}_i^{ref}(t) = -k_i ACE_i(t), \tag{5}$$

where k_i is the integrator gain of the PI-controller, and ACE_i is the area control error in area i, computed as

$$ACE_i(t) = \Delta P_i^{tie}(t) + \beta_i\Delta\omega_i(t), \tag{6}$$

where β_i is the frequency bias of area i computed as

$$\beta_i = D_i + \frac{1}{R_i}. \tag{7}$$

A block diagram of the above equations for two areas is shown in Figure 1, where the transfer function of each block is given in the Laplace domain.

The above equations can be converted into the state space model

$$\dot{x}_i(t) = A_{ii}^c x_i(t) + B_i^c u_i(t) + \sum_{j=1}^N A_{ij}x_j(t), \tag{8}$$

where

$$x_i = \left[\Delta P_i^{tie},\ \Delta\omega_i,\ \Delta P_i^m,\ \Delta P_i^v,\ \Delta P_i^{ref}\right]^T,\ u_i = \Delta P_i^L,$$

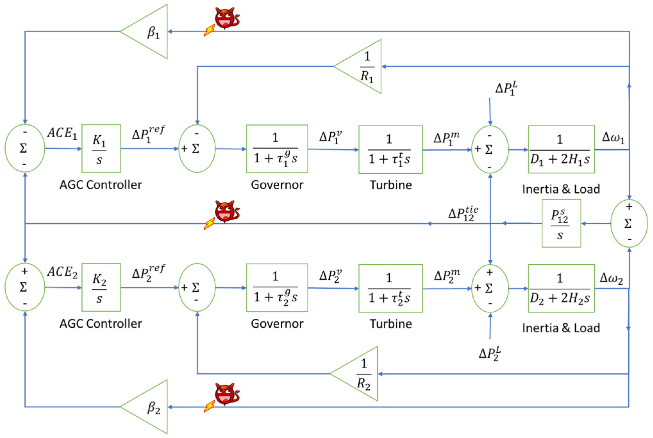

Fig. 1. Block diagram of automatic generation control of a 2-area power system using ACE, including the locations of FDIAs.

$$A_{ii}^c = \begin{bmatrix} 0 & \sum_{j=1}^N P_{ij}^s & 0 & 0 & 0 \\ \frac{-1}{2H_i} & \frac{-D_i}{2H_i} & \frac{1}{2H_i} & 0 & 0 \\ 0 & 0 & \frac{-1}{\tau_i^t} & \frac{1}{\tau_i^t} & 0 \\ 0 & \frac{-1}{R_i\tau_i^g} & 0 & \frac{-1}{\tau_i^g} & \frac{1}{\tau_i^g} \\ -k_i & -\,k_i\beta_i & 0 & 0 & 0 \end{bmatrix},$$

$$B_i^c = \left[0,\ \frac{-1}{2H_i},\ 0,\ 0,\ 0\right]^T,$$

and A_{ij}^c, $i \neq j$ is a 5×5 matrix whose only non-zero element is $-P_{ij}^s$ in the first row and the second column.

Combining the equations for all areas we obtain

$$\dot{x}(t) = A^c x(t) + B^c u(t), \tag{9}$$

where $x \in \mathbb{R}^{5N}$, $u \in \mathbb{R}^N$, $A \in \mathbb{R}^{5N\times 5N}$, $B \in \mathbb{R}^{5N\times N}$ s.t.

$$x = \left[x_1^T, \ldots, x_N^T\right]^T,\ u = \left[u_1^T, \ldots, u_N^T\right]^T,$$

$$A^c = \begin{bmatrix} A_{11}^c & \ldots & A_{1N}^c \\ \vdots & \ddots & \vdots \\ A_{N1}^c & \ldots & A_{NN}^c \end{bmatrix},\ B^c = \begin{bmatrix} B_1^c & 0 & \ldots & 0 \\ 0 & B_2^c & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & B_N^c \end{bmatrix}.$$

The above continuous time model can be converted to discrete time with a discretization time step T_s using the zero-order hold (ZOH) method [47] to obtain

$$x[t+1] = Ax[t] + Bu[t], \tag{10}$$

where A and B are obtained by the ZOH discretization of A^c and B^c respectively.

B. Fault and Attack Detection in AGC

As mentioned in Section II, the most commonly used methods for fault and attack detection in AGC are (1) monitoring the ACE values, which is a model-free method, and (2) developing unknown input observers, which is model-based.

1) Area Control Error: The ACE can be computed for each area as in (6), based on the received power-flow and frequency measurements. Since the main objective of AGC is to keep the ACE values small, an increase in ACE could be a strong indicator for a system fault or malicious activity [9]. The simplest

ACE-based detector would then monitor the ACE values, and raise an alarm if

$$\max_i |ACE_i[t]| > \rho_a, \tag{11}$$

where ρ_a is a predefined detection threshold. In what follows, we refer to the detector based on (11) as the *ACE detector*.

2) Unknown Input Observer: Another commonly used method for fault and attack detection for AGC is based on the idea of the delayed unknown input observer (UIO) for discrete-time linear systems [4], [19]. The UIO is based on the discrete-time state space model of the system,

$$\begin{aligned} x[t+1] &= Ax[t] + Bu[t] \\ y[t] &= Cx[t], \end{aligned} \tag{12}$$

where $y \in \mathbb{R}^{3N}$, $C \in \mathbb{R}^{3N \times 5N}$ s.t.

$$y = \left[y_1, \ldots, y_N\right]^T, \ C = \begin{bmatrix} C_1 & 0 & \ldots & 0 \\ 0 & C_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & C_N \end{bmatrix},$$

$$y_i = \left[\Delta P_i^{tie}, \ \Delta\omega_i, \ \Delta P_i^{ref}\right], \ C_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that the output y includes only the variables that could be measured by the operator. For ease of notation we let $n = 5N$, $m = N$, $p = 3N$ denote the total number of states, inputs, and outputs of the system, respectively. Assuming knowledge of the initial state of the system (i.e., $x[0]$), a UIO with a detection delay α can be used to estimate the system state at time t after observing the system measurements y from time t to $t+\alpha$, making use of the relation

$$y[t : t+\alpha] = \Theta_\alpha x[t] + M_\alpha u[t : t+\alpha], \tag{13}$$

where $y[t : t+\alpha] \in \mathbb{R}^{p(\alpha+1)}$, $u[t : t+\alpha] \in \mathbb{R}^{m(\alpha+1)}$, $\Theta_\alpha \in \mathbb{R}^{p(\alpha+1)\times n}$, $M_\alpha \in \mathbb{R}^{p(\alpha+1)\times m(\alpha+1)}$ s.t.

$$\begin{aligned} y[t : t+\alpha] &= \left[y[t]^T, y[t+1]^T, \ldots, y[t+\alpha]^T\right]^T, \\ u[t : t+\alpha] &= \left[u[t]^T, u[t+1]^T, \ldots, u[t+\alpha]^T\right]^T, \\ \Theta_\alpha &= \left[C^T, (CA)^T, \ldots, (CA^\alpha)^T\right]^T, \\ M_\alpha &= \begin{bmatrix} 0 & 0 & \ldots & 0 \\ CB & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\alpha-1}B & CA^{\alpha-2}B & \ldots & 0 \end{bmatrix}, \end{aligned}$$

The estimated system state $\hat{x}[t]$ by the UIO can then be given by

$$\hat{x}[t+1] = A\hat{x}[t] + L\big(y[t : t+\alpha] - \Theta_\alpha \hat{x}[t]\big), \tag{14}$$

where $L \in \mathbb{R}^{n\times p(\alpha+1)}$ is the UIO gain matrix that should be designed in order to ensure the accuracy and stability of the UIO. It has been shown [19], [48] that for $\alpha \geq 2$, the accuracy and stability of the UIO can be ensured when the following procedure is followed [4], [19]:

1) Choose α s.t. $rank(M_\alpha) - rank(M_{\alpha-1}) = N$
2) Find the matrix $Q \in \mathbb{R}^{\alpha m\times(\alpha+1)p}$ that satisfies
$$QM_\alpha = \begin{bmatrix} 0 & 0 \\ I_m & 0 \end{bmatrix}$$
3) Compute $[S_1^T \quad S_2^T]^T = Q\Theta_\alpha$ s.t. $S_1 \in \mathbb{R}^{(\alpha-1)m\times n}$ and $S_2 \in \mathbb{R}^{m\times n}$
4) Find $L_1 \in \mathbb{R}^{n\times(\alpha-1)m}$ s.t. the eigenvalues of $(A - BS_2) - L_1 S_1$ are stable (i.e., $\in [-1, 1]$)
5) Compute $L = [L_1 \quad B] \ \times Q$

After estimating the system state $\hat{x}$ using (14), the residual can be computed as

$$r[t] = y[t] - C\hat{x}[t], \tag{15}$$

and an alarm is raised if

$$\|r[t]\|_2 > \rho_r, \tag{16}$$

where ρ_r is a predefined detection threshold. Recall that despite being the residual of the estimated state for time t, $r[t]$ cannot be computed by the UIO before time $t+\alpha$. Furthermore, observe that the knowledge of the load changes (i.e., $u[t]$) is not required to compute $r[t]$. In what follows we refer to the detector based on (16) as the *UIO detector*.

C. Attack Model

We consider an attacker that has knowledge of the system matrices A, B, and C. This means that the attacker either knows or can accurately estimate the parameters of each area (i.e., $H_i, D_i, R_i, \beta_i, \tau_i^g, \tau_i^t, P_{ij}^s$). Furthermore, the attacker is able to eavesdrop on the system measurements (i.e., $y[t]$) at each AGC cycle. We assume that the attacker knows whether the system operator is using an ACE detector, a UIO detector, none, or both. If the operator is using a UIO, the attacker knows the parameters α and L of the UIO, and can thus predict the effect of its attack on the UIO residual r.

We consider that the attacker can inject false measurements of the tie-line power flows as well as the area frequencies, and can thus manipulate ΔP^{tie} and $\Delta\omega$ as

$$\begin{aligned} \Delta P_{i,a}^{tie}[t] &= \Delta P_i^{tie}[t] + a_i^P[t], \\ \Delta\omega_{i,a}[t] &= \Delta\omega_i[t] + a_i^\omega[t], \end{aligned} \tag{17}$$

where a_i^P and a_i^ω represent the perturbation (attack) of the tie-line power flows and frequency in area i. Observe that in practice, manipulating the frequency measurements might be harder than manipulating power flow measurements since (1) power flow measurements are typically greater in number than frequency measurements, and are thus harder to secure, and (2) the grid frequency is a variable that can be verified by the system operator from neighbouring buses in the same area [3]. The attacked power-flow and frequency measurements would then affect the ACE computation in (6), and thus the output of the PI-controller in (5).

In practice, the attacker could eavesdrop and inject false measurements through network intrusion. The attack could directly manipulate messages transmitted using communication protocols such as Modbus, DNP3 or IEC 61850 [49], [50], as these protocols do not mandate either authentication or encryption of messages. Although security recommendations exist for these protocols [51], their use is not mandatory. Even if message authentication is used, eavesdropping and injection of measurements would be feasible through the compromise of end devices. An end device (e.g., a remote terminal unit (RTU) or a phasor measurement unit (PMU)) could be compromised by stealing cryptographic credentials or by exploiting software or hardware vulnerabilities, and state estimates based on PMU measurements could also be compromised by time synchronization attacks [52]. Finally, information regarding the system parameters and the used detectors could be obtained by insiders, or could be estimated by an adversary that can eavesdrop the measurements during an extended period of time.

Overall, the advantage of such a strong attack model is that it allows us to consider the worst case attacks and their potential impact on the system's performance. Such a strong model is not uncommon in the security literature, given the recent success of cyber attacks with high level of attacker knowledge, e.g., Stuxnet [53] and FDIAs against power system state estimation [54], [55]. We further assume that the attack is constrained by

$$\begin{aligned} \left|\Delta P^{tie}_{i,a}[t]\right| &\leq a^{P+}, \\ \left|\Delta\omega_{i,a}[t]\right| &\leq a^{\omega+}, \end{aligned} \tag{18}$$

where a^{P+} and $a^{\omega+}$ denote the respective maximum allowed attack magnitudes. The reason for constraint (18) is that an attack that sets the power-flow or frequency measurements too far from their expected values should be easily detectable. Moreover, the attacker is constrained by that $\sum_{i=1}^{N}\Delta P^{tie}_{i,a}$ and $\sum_{i=1}^{N} a^P_i$ must be kept close to zero for any attack. As a result of the attack, the state-space model becomes [20]

$$x'[t+1] = Ax'[t] + Bu[t] + Ea[t], \tag{19}$$

$$y'[t] = Cx'[t] + Fa[t], \tag{20}$$

where $a \in \mathbb{R}^{2N}$. $E \in \mathbb{R}^{n\times 2N}$, $F \in \mathbb{R}^{p\times 2N}$ s.t.

$$a = \left[a^P_1, a^\omega_1, \dots, a^P_N, a^\omega_N\right]^T, \tag{21}$$

$$E = \begin{bmatrix} E_1 & 0 & \dots & 0 \\ 0 & E_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & E_N \end{bmatrix},$$

$$F = \begin{bmatrix} F_1 & 0 & \dots & 0 \\ 0 & F_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & F_N \end{bmatrix}, \quad E_i = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -k_i & -k_i\beta_i \end{bmatrix},$$

$$F_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Observe that the only state variable that is directly affected by the attack is ΔP^{ref}, due to the manipulated ACE value. The considered attack model is illustrated in Figure 2.

The attacker's goal is to maximize the deviation of the frequency from its nominal value f_0 in a certain target area i^*. Ideally, the attacker would like to cause the frequency to drift beyond its secure limit, which might cause load shedding schemes to take effect, or in the worst case cause blackouts.

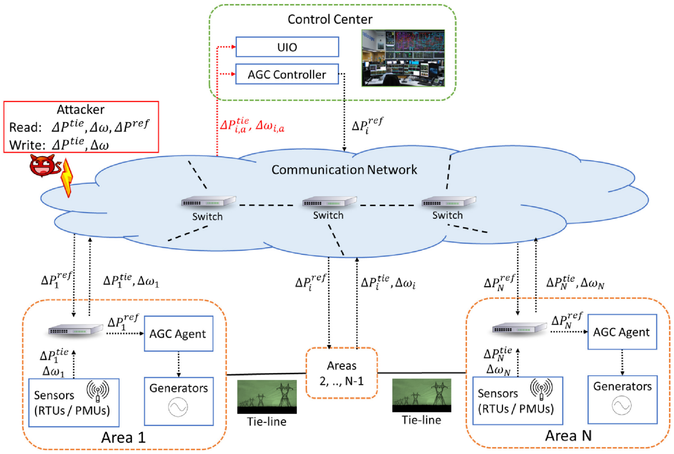

Fig. 2. Block diagram of the AGC system, including the physical power system, the communication network, the control center, and the attacker.

We consider that the adversary aims to find a sequence $\pi = (a[1], a[2], \dots, a[T])$ for some time horizon T that maximizes the frequency deviation without being detected by either the UIO or the ACE. This corresponds to solving the optimization problem

$$\begin{aligned} \max_{\pi}\ & \frac{1}{T}\sum_{t=1}^{T}|\Delta\omega_{i^*}[t]|, \\ \text{s.t.}\ & \|r[t]\|_2 \leq \rho_r, \quad t = 1, 2, \dots, T \\ & \max_i\ |\mathrm{ACE}_i[t]| \leq \rho_a, \quad t = 1, 2, \dots, T. \end{aligned} \tag{22}$$

An important feature of this seemingly simple problem is that the attacker has limited information about the system at every time step and has no knowledge of the future evolution of the system. Thus, (22) is essentially a sequential decision problem under uncertainty, and hence we propose to adopt a multi-objective POMDP formulation.

IV. RL-Based Attacks on AGC

In what follows we formulate the problem of computing attacks against AGC that are undetectable w.r.t. the UIO and the ACE as a multi-objective partially observable Markov decision process (MO-POMDP) [5], and propose to use reinforcement learning for obtaining an attack policy. Although we present a solution that specifically targets the two detectors discussed in Section III-B, the proposed approach can easily be extended to target any other model-based or model-free fault and attack detection method that is based on a hypothesis test in the form (11) or (16).

A. Multi-Objective POMDP Formulation

We formulate the problem by first introducing a tuple M and then showing that it is a POMDP. Let $M \triangleq (\mathcal{S}, \mathcal{A}, R, \mathcal{P}, \mathcal{O}, \gamma)$, where:

- $\mathcal{S}$ is the state space, and $s[t] \in \mathcal{S}$ is the state at time step t. For our problem, this includes the state of the AGC system, the load demand, the current estimated state by the UIO, as well as the delayed measurements needed for estimating the next state. Therefore,

$$s[t] \triangleq \big(x\big[t\big], u[t], \hat{x}[t-\alpha], y'[t-\alpha-1 : t-1]\big). \quad (23)$$

- $\mathcal{A}$ is the set of the attacker's possible actions, and $a[t] \in \mathcal{A}$ denotes the action at time step t as defined in (21).
- $R[t]$ is the reward function. We propose a reward function that rewards an increase of the frequency deviation at the target area and at the same time includes punishment terms for the UIO residual and the ACE. Particularly, we use a weighted sum of the frequency deviation, the norm of the residual, and the maximum of the ACE values among different areas as the reward,

$$\begin{aligned} R[t] \triangleq\ & |\Delta\omega_{i^*}[t+1]| \\ & - \lambda_r \|r[t-\alpha+1]\|_2 - \lambda_a \max_i |\mathrm{ACE}_i[t]|, \quad (24) \end{aligned}$$

where λ_r and λ_a are regularization coefficients (note that $\Delta\omega_{i^*}[t+1]$, $r[t-\alpha+1]$, and $\mathrm{ACE}_i[t]$ are the resulting frequency deviation, residual, and ACE, when the transition $(s[t], a[t]) \rightarrow s[t+1]$ occurs). The values of (λ_r, λ_a) can be used for setting the relative importance of the impact ($\Delta\omega_{i^*}$) and (un)detectability (r and ACE). Observe that the reward function essentially converts three objectives into a scalar objective, which is a widely used approach for dealing with MO-POMDPs.
- $\mathcal{P}(s[t+1]|s[t], a[t])$ represents the conditional transition probability between states.
- $\mathcal{O}$ denotes the attacker's observation space. At each time step t the attacker obtains the observation $o[t] \in \mathcal{O}$ about the state. We define this observation as

$$o[t] \triangleq \big(y\big[t\big], y'[t-\alpha-1 : t-1]\big). \quad (25)$$

Note that we assumed that the vector y is observable by the attacker, and y' is the result of the attacker's actions on y and accordingly, is observable by the attacker.
- $\gamma \in [0, 1)$ is a discount factor.

Proposition 1: The tuple M with the definitions in (23), (24), and (25) is a POMDP.

Proof: To prove this, we need to show that

(i) $s[t]$ as defined in (23) is indeed Markovian.

(ii) The transition $(s[t], a[t]) \rightarrow s[t+1]$ contains all information needed for computing the reward.

In order to prove (i), we need to show that $s[t+1]$ only depends on $s[t]$ and $a[t]$ and not the entire history, i.e., we have to verify that

$$\begin{aligned} & p\big(s[t+1]\big|\ s[t], a[t]\big) \\ & \quad = p\big(s[t+1]\big|\ s[t], s[t-1], \ldots, s[0], a[t]\big). \quad (26) \end{aligned}$$

Equations (14), (19), and (20) show that $\hat{x}[t-\alpha+1]$, $x[t+1]$, and $y'[t]$ (and accordingly, $y'[t-\alpha : t]$) are independent of $s[t-1], \ldots, s[0]$ given $s[t]$ and $a[t]$. Assuming that $u[t]$ has the Markov property (e.g., a random walk), then the state $s[t]$ has the Markov property as well.

Regarding (ii), notice that the term $\Delta\omega_{i^*}[t+1]$ is an entry of $x[t+1]$, which itself is a part of $s[t+1]$. In addition, $r[t-\alpha+1]$ can be obtained as

$$r[t-\alpha+1] = y'[t-\alpha+1] - C\hat{x}[t-\alpha+1],$$

and both $y'[t-\alpha+1]$ and $\hat{x}[t-\alpha+1]$ are included in $s[t+1]$. Finally, $\mathrm{ACE}_i[t]$ is computed based on $y'[t]$ using (6), and $y'[t]$ is determined by $s[t]$ and $a[t]$. Hence, writing the reward in (24) as $R[t] = R(s[t], a[t], s[t+1])$ is well-justified. ■

B. Attacker's Policy

To solve the above-mentioned POMDP, the attacker seeks to find a policy $\pi : \mathcal{O} \rightarrow \mathcal{A}$ that maximizes the expected discounted average reward. That is, the attacker's objective is finding the solution to the following problem:

$$arg\,max_{\pi}\ \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R[t]\right] \quad (27)$$

Note that maximizing the objective in (27) corresponds to solving the following optimization problem:

$$\begin{aligned} & \max_{\pi}\ \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R[t]\right] \\ & = \max_{\pi}\ \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \Big(|\Delta\omega_{i^*}[t+1]| - \lambda_r \|r[t-\alpha+1]\|_2 \right. \\ & \qquad \left. - \lambda_a \max_i |\mathrm{ACE}_i[t]|\Big)\right] \\ & = \max_{\pi}\ \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t |\Delta\omega_{i^*}[t+1]|\right] \\ & \quad - \lambda_r \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \|r[t-\alpha+1]\|_2\right] \\ & \quad - \lambda_a \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \max_i |ACE_i[t]|\right], \end{aligned}$$

which can be regarded as a relaxed approximation of the problem in (22). This justifies our definition of the reward function in (24).

Finding the optimal policy is an RL problem with continuous state and action spaces. We thus propose to use deep RL for finding good policies. In what follows we refer to the attack based on this policy as the deep RL attack DRLA(λ_r, λ_a).

V. Numerical Results

In this section we evaluate the proposed RL-based attacks and compare them to state-of-the-art FDIAs against AGC. All experiments were carried out on a server with AMD 7543P CPU with 32 cores @ 2.8 GHz and 64 GB of RAM.

A. Simulation Methodology

We simulated an $N = 3$-area power system operating at a nominal frequency of 60 Hz. The parameters for areas 1 and 2 are the same as for the examples in [2, Ch. 12] and the parameters for area 3 were obtained by slightly perturbing the values for area 1, as shown in Table I. Each area is connected to the other two areas through a tie line. Although seemingly simplistic, the simulated 3-area system can model a wide-range of practical systems, since each area does in reality include many generator and load buses. To simulate the dynamics of

TABLE I
PARAMETERS OF THE CONSIDERED THREE-AREA POWER SYSTEM

Parameter	Area 1	Area 2	Area 3
H_i	5	4	5.5
D_i	0.6	0.9	0.7
τ_i^t	0.5	0.6	0.51
τ_i^g	0.2	0.3	0.25
R_i	0.05	0.0625	0.0525
K_i	0.3	0.3	0.3
BaseMVA	1000	1000	1000
P_i^s	2	2	2

the system, we assumed a discretization time step of $T_s = 2$ seconds, which is a reasonable value considering the AGC cycle [2]. The load for each area is assumed to follow a random walk given by

$$\Delta P_i^L[t+1] = \Delta P_i^L[t] + v_i^L[t], \tag{28}$$

where v_i^L follows a zero-mean Gaussian distribution with a standard deviation $\sigma_i^L = 0.02$ p.u. for all areas. Furthermore, state noise and measurement noise are added to (12) according to zero-mean Gaussian distributions with a standard deviation of 0.03 Hz for frequency variables and $\sqrt{0.03}$ MW for power variables [19], [56]. The above three factors (i.e., load fluctuation, state and measurement noise) are thus the main sources of randomness in our experiments. For the evaluation we implemented a UIO with an estimation delay of $\alpha = 2$, which is the smallest value that ensures the accuracy and stability of the UIO [48]. To choose the UIO gain matrix L, the eigenvalues of the matrix $(A - BS_2) - L_1S_1$ were chosen to be equidistant values in the range $[-0.5, 0.5]$, which was observed to improve the UIO accuracy.

Next, to compute DRLAs, we considered that the maximum allowed deviation of the power-flow is $a^{P+} = 0.3$ p.u. $= 300$ MW, and the maximum allowed deviation of the frequency (in the case frequency measurements are attacked) is $a^{\omega+} = 0.006$ p.u. $= 0.36$ Hz, as in (18). The aforementioned values were chosen based on preliminary experiments s.t. the deviations in attacked measurements are large enough to affect the AGC system, but not too large to raise alarms and initiate remedial actions by the operator. The attack objective was to maximize the frequency deviation in area 1, i.e., $i^* = 1$. Since the states and actions are continuous, popular discrete-space RL algorithms such as deep Q-network (DQN) [57] could not be used. Instead, the RL attacks were trained by the proximal policy optimization (PPO) algorithm [58]. PPO was chosen based on the results of preliminary experiments comparing its performance to other state-of-the-art continuous-space RL algorithms such as deep deterministic policy gradient (DDPG) [59], and soft actor-critic (SAC) [60]. Due to its simplicity, ease of tuning, and state-of-the-art performance in various RL tasks, PPO is currently one of the most used RL algorithms. It belongs to the class of actor-critic policy gradient algorithms. The PPO algorithm consists of two interacting neural networks: an actor network which learns to produce actions based on observations, and a critic network which learns to evaluate the actions generated by the actor network. The actions produced by the actor NN are optimized by maximizing the clipped value of the advantage function, which quantifies the advantage of taking an action compared to the average behavior. The optimization objective could possibly include minimizing the KL-divergence [61] between the policies followed in subsequent optimization steps. In our PPO implementation, we used the default PPO parameters from the RL-lib Python library [62]. The discount factor used was $\gamma = 0.99$. The advantage function was estimated using generalized advantage estimation (GAE) [63] with $\lambda_{GAE} = 1$. The KL-divergence was included in the objective with a coefficient of 0.2 and a target of 0.01. The PPO clip parameter used was $\epsilon = 0.3$. The actor and critic NNs were implemented in the Tensorflow Python library [64], and each network included 2 hidden layers with 256 neurons, and *tanh* activation functions. The NNs were optimized using stochastic gradient descent (SGD) [65] with 30 epochs of training per batch, and a mini-batch size of 128 samples. The number of episodes needed to train each RL agent was 80,000. Each episode's length was 150 AGC cycles (i.e., 300 seconds given $T_s = 2$s), and the attacks started at the 51st cycle, resulting in $T_e = 100$ attacked AGC cycles per episode. The initial 50 unattacked cycles were simulated to avoid any undesired interaction between the attack and the initial transient behavior of the UIO. We used three attack schemes as baselines for comparison.

a) *Random Attack:* At each time step, the attack is randomly chosen according to a uniform distribution, i.e., $\Delta P_{i,a}^{tie} \sim \mathcal{U}(-a^{P+}, a^{P+})$ and $\Delta\omega_{i,a} \sim \mathcal{U}(-a^{\omega+}, a^{\omega+})$.

b) *Regression Attack:* proposed in [3], the attacker develops a linear regression model of the attack impact (i.e., $|\Delta\omega_1|$) as a function of the change in the area loads $\Delta P^L[t]$, and the attacker's action $a[t]$. The optimal attack can then be computed based on the learned model.

c) *DRLA* $(\mathbf{0}, \mathbf{0})$*:* the attacker attempts to maximize the impact, without taking neither the UIO residual nor the ACE into consideration, and uses RL for this purpose. This is achieved by setting $\lambda_r = \lambda_a = 0$ in (24).

For each attack scenario, the simulation procedure is as follows at each time step:

1) Compute the attack $a[t]$ according to the attack policy.
2) Compute the attacked measurements $y'[t]$ as in (20).
3) Compute the UIO residuals r based on $y'[t]$.
4) Simulate the state-space model of the AGC system according to (19).
5) Compute the un-attacked measurements $y[t+1]$ from (12), which will be part of the observation $o[t+1]$ for the attacker.

B. Attack Impact and Detectability

In what follows we present the results of the evaluation of our proposed DRLA s against AGC. Figure 3(left) and (right) shows the attack impact measured as the maximum frequency deviation in the target area (i.e., Area 1) during an episode vs. the maximum ℓ_2-norm of the UIO residual over one episode, and the attack impact vs. the maximum $\max_i |ACE_i|$ over one episode, respectively, when attacking only the power-flow

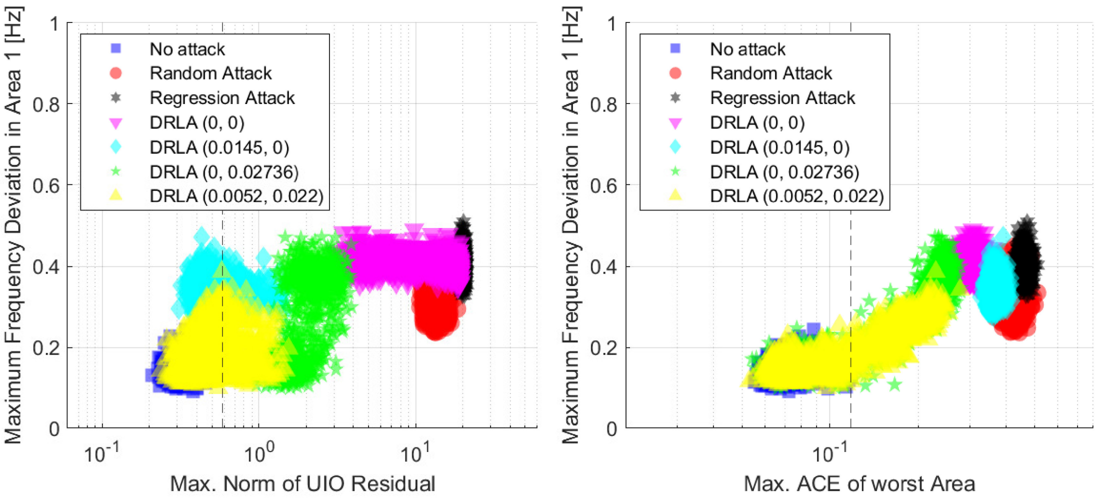

Fig. 3. Trade-off between attack impact and detection metrics for DRLAs and baselines, when only power-flow measurements are attacked.

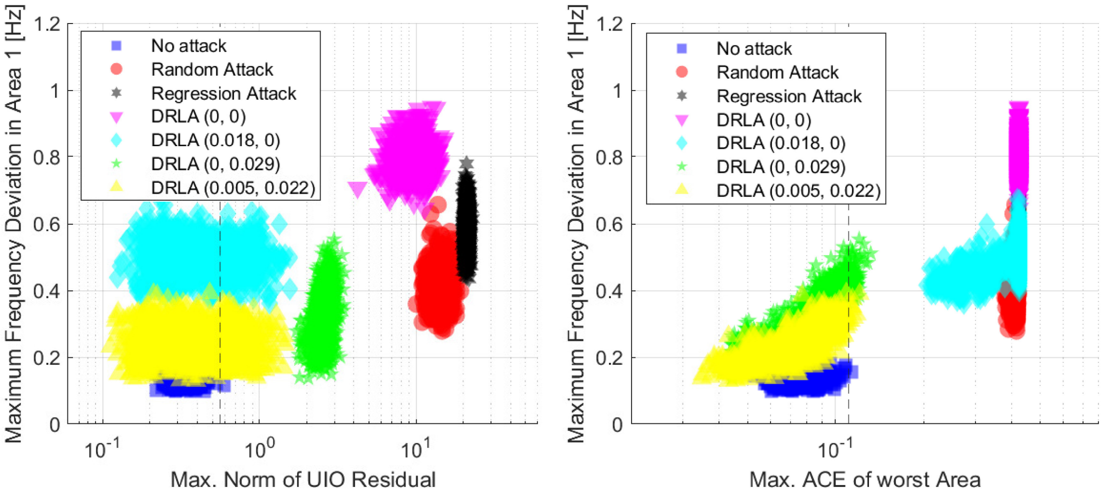

Fig. 4. Trade-off between attack impact and detection metrics for DRLAs and baselines, when both power-flow and frequency measurements are attacked.

measurements with ($a^{P+} = 0.3$). Figure 4 shows corresponding results for the case when attacking both power-flow and frequency measurements with ($a^{P+} = 0.3, a^{\omega+} = 0.006$). Each point in the figures represents one episode and a total of 1000 episodes were simulated per scenario. We identified non-zero λ_r and λ_a values by numerically exploring the Pareto frontier, and then choosing parameter pairs with significant impact while retaining undetectability. Focusing on Figure 3, we can first observe that the baselines can typically achieve slightly higher impact compared to the proposed DRLA s. For example, the maximum impact achieved by the baselines is around 0.5 Hz, compared to slightly over 0.4 Hz for DRLA s. However, DRLA s (with non-zero regularization coefficients) can greatly reduce the values of the detection metrics compared to baselines (e.g., by around two orders of magnitude for the UIO residual and around one order of magnitude for the ACE), and bring the detection metrics close to their values in the no-attack scenario. Furthermore, as expected, DRLA(0.0145, 0) which penalizes high UIO residuals succeeds in achieving a good balance between impact and UIO residuals. However, it clearly fails in keeping the ACE values low (similar to the baselines). The exact opposite is observed for DRLA(0, 0.02736). Comparing the two aforementioned attacks, it can be observed that attacking the UIO residual seems to be easier than attacking the ACE, which indicates that the ACE might be a better metric for detecting attacks than the UIO residuals in this case. On the contrary, DRLA(0.0052, 0.022) succeeds in keeping both detection metrics low, at the cost of lower attack impact.

Comparing the above results with Figure 4, we can observe that attacking the frequency measurements can allow the attacker to slightly increase both the attack impact and stealthiness. For example, DRLA(0.018, 0) can have an impact reaching 0.6 Hz, which is above the security limit of many applications. Furthermore, the same attack yields UIO residuals that are on average much lower than the corresponding attack in Figure 3. Note the discrepancy between the values of (λ_r, λ_a) in Figures 3 and 4, since these values were chosen empirically. The vertical line in the figures shows the detection threshold corresponding to a false positive rate (FPR) of 0.1%. The FPR is defined as the fraction of non-attacked episodes for which the detector raises an alarm, and can be controlled

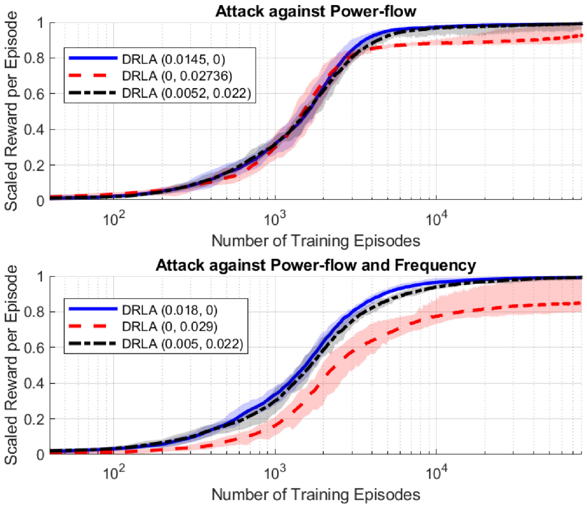

Fig. 5. Reward curves during training.

by changing the detection threshold, and FPR=0.1% corresponds to a time between false alarms TBFA $= T_e T_s/\text{FPR} = 100 \times 2/0.001 = 200,000$ seconds ($<$ 0.5 false alarms per day). Figure 4 shows that the *UIO detector* can detect 27.6% of the DRLA(0.005, 0.022) attacks and 28.2% of the DRLA (0.018, 0) attacks. For the same FPR, the *ACE detector* can detect only 1.1% of the DRLA(0.005, 0.022) attacks and 8.3% of the DRLA(0, 0.029) attacks. This suggests that the *UIO detector* is better than the *ACE detector* for this case. In general, Figure 4 confirms the earlier observation that DRLA is successful in terms of impact and (un)detectability. We have also evaluated the performance of an additional detector based the cumulative sum (CUSUM) [66] of the UIO residuals. The results, shown in the Appendix, suggest that CUSUM does not provide a significant improvement over the above detectors, especially for the case when both power-flow and frequency measurements are attacked.

C. Training Stability

To assess the stability of DRLA, we further trained 10 separate agents for each (λ_r, λ_a) tuple, excluding the baseline DRLA(0, 0), and computed the minimum, mean, and maximum reward per episode over the 10 agents as the training progresses. Figure 5 shows the so-called reward curves for the trained agents, with and without attacking frequency measurements. To facilitate the comparison, the rewards were scaled over the 10 trained agents using min-max scaling. The figure shows that most agents do converge with very low variance after around 10,000 episodes of training, with the only exception being DRLA(0, 0.029) (when attacking both ΔP^{tie} and $\Delta\omega$), which indicates that the agents might need further training. To conclude, the trained agents show in general very stable performance.

D. Immediate Response

We further consider the hypothetical scenario that the operator immediately reacts to the attacks detected by either the UIO or the ACE detectors (e.g., through neglecting suspected measurements, or initiating load shedding schemes). For this case, it is reasonable to evaluate the attacks in terms of the highest impact caused until detection, instead of the highest impact over the whole episode. For brevity, all upcoming results concern the scenario where both power-flow and frequency measurements are attacked (i.e., $a^{P+} = 0.3,\ a^{\omega+} = 0.006$), unless otherwise stated. Figure 6 shows the relation between the attack impact before detection, and the average TBFA. Every point is computed by using a different value for the detection thresholds (ρ_r or ρ_a). The figure shows that the effective impact of the baseline attacks is always negligible irrespective of the chosen TBFA, since those attacks are always detected at the beginning of an episode, before they can achieve any significant impact. Interestingly, this is also the case for DRLA s targeting the wrong detection metrics, e.g., DRLA s with $\lambda_r = 0$ have negligible effective impact when the UIO detector is used, and vice versa. Among DRLA s, the effective impact of the attacks with non-zero regularization coefficients increases with the TBFA until it approaches the average impact shown in Figure 4. The results in this figure and the previous figures emphasize the importance of the attacker's knowledge of the detector employed by the defender. They also show that even if the operator decides to use both detection metrics, DRLA s with $\lambda_r > 0$ and $\lambda_a > 0$ are expected to be undetected, even if somewhat less impactful.

E. Data-Driven Detectors

To further investigate the detectability of the proposed DRLA s, we examine the use of two machine learning (ML) based detection approaches: (1) an unsupervised autoencoder (AE) neural network, and (2) a supervised deep neural network (DNN) classifier. For both approaches, we consider that the input features at each timestep are: (a) the measurements $y[t]$, (b) the UIO residuals $r[t-\alpha]$, (c) the norm of the UIO residuals $\|r[t-\alpha]\|_2$, and (d) the ACE in all areas. Thus, for our 3-area system this corresponds to a total of $n_f = 9+9+1+3 = 22$ features. The dimensions of the AE layers were $n_f * [1,\ 0.7,\ 0.5,\ 0.7,\ 1]$ (i.e., three hidden layers), and the dimensions of the DNN layers were $n_f * [1,\ 4,\ 0.5,\ 1]$ (i.e., two hidden layers). Both approaches used ReLU as the activation function for the neurons, used the Adam optimization algorithm, and were implemented using PyTorch. To evaluate the data-driven detectors, we used the same simulation data described in Section V-B. The data (7 attack scenarios × 1000 episodes × 100 time steps) were split into 800 training episodes and 200 test episodes. The unsupervised AE was trained on non-attacked training data only, while the supervised DNN was trained using the whole labelled training data. The detection was then done on the test data using a hypothesis test similar to (11) and (16), where the test statistics for AE and DNN were the MSE of the AE reconstruction error (the difference between input and output layers), and the scalar output of the DNN, respectively.

To compare the performance of the ML detectors to the UIO and ACE detectors, we utilize the receiver operating characteristic (ROC) curves. The ROC curve shows the trade-off between the fraction of attacked episodes for which a detector

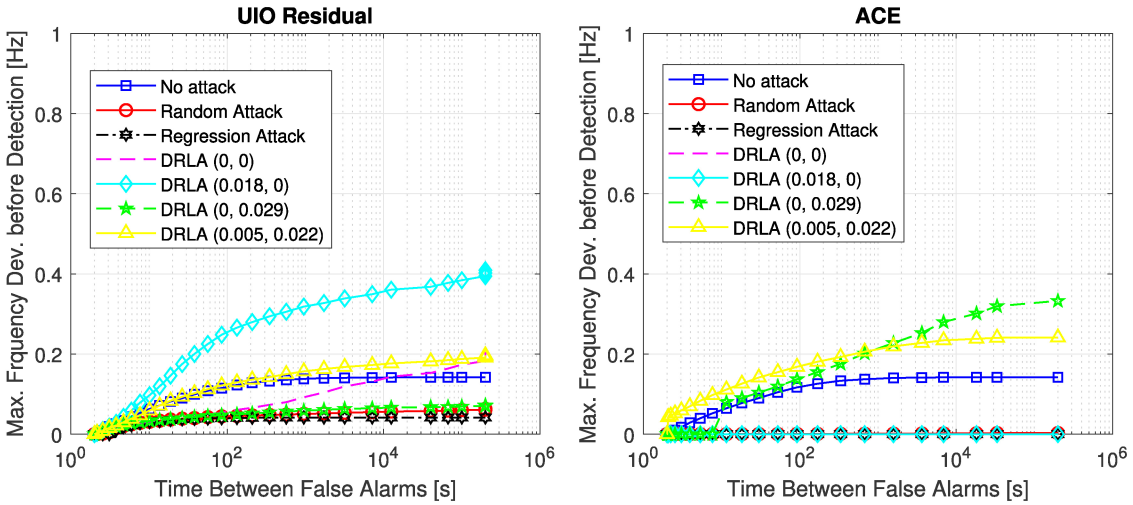

Fig. 6. Trade-off between the highest achieved attack impact before detection and the time between false alarms.

TABLE II
COMPARISON OF THE ATTACKS W.R.T. THEIR IMPACT AND CORRESPONDING AUC SCORES BY THE DIFFERENT DETECTORS

Scenario	Impact	UIO	ACE	AE	DNN
No attack	0.1422	-	-	-	-
Random attack	0.4056	1	1	1	1
Regression attack	0.5857	1	1	1	1
DRLA$(0, 0)$	0.7947	1	1	1	1
DRLA$(0.018, 0)$	0.4987	0.5694	1	1	1
DRLA$(0, 0.029)$	0.3412	1	0.6743	1	1
DRLA$(0.005, 0.022)$	0.2417	0.5610	0.5388	0.7753	0.9995

TABLE III
COMPARISON OF THE ATTACK IMPACTS AND CORRESPONDING AUC SCORES, IN THE PRESENCE OF 20% PARAMETER MISESTIMATION

Scenario	Impact	UIO	ACE	AE	DNN
No attack	0.2270	-	-	-	-
Random attack	0.6084	1	1	1	1
Regression attack	0.8300	1	1	1	1
DRLA$(0, 0)$	1.1280	1	1	1	1
DRLA$(0.018, 0)$	0.5917	0.3670	1	1	1
DRLA$(0, 0.029)$	0.4566	1	0.1317	1	0.9997
DRLA$(0.005, 0.022)$	0.3445	0.3841	0.0514	0.6531	0.9989

raises an alarm (true positive rate, or TPR) on the vertical axis, and the FPR on the horizontal axis, and is obtained by varying the detection threshold (e.g., ρ_r in (16) for the UIO detector). The area under the ROC curve is a commonly-used evaluation metric that summarizes the performance of the detector. An ideal detector would have AUC = 1, while a detector with AUC = 0.5 would correspond to a performance that is as good as random guessing.

Table II shows the AUC achieved by each of the detectors, as well as the mean impact of each attack. Observe that the impact was defined as the maximum observed frequency deviation in area 1, and hence the no-attack scenario has non-zero impact. From the table, we can generally see that the ML detectors (especially the DNN) have significantly higher AUC values than the UIO and ACE detectors. Nonetheless, the unsupervised AE performs surprisingly poor against DRLA (0.005, 0.022), even though the attack was not specifically trained to bypass it. This result interestingly indicates the potential generalization power of DRLA against unseen detectors. Finally, it is worth noting that although the supervised DNN can effectively detect all considered attacks, the performance of supervised ML typically degrades against unseen and zero-day attacks [67], [68]. Moreover, the acquisition of accurately labelled data in real scenarios might not always be feasible [69], [70].

F. Impact of Parameter Misestimation

We now consider the case when the operator's model of the AGC system (i.e., H, D, τ^t, τ^g, R) is slightly inaccurate. Model inaccuracy would affect the control accuracy of the PI-controllers and the UIO residuals, making an attack potentially more difficult to detect. To simulate this scenario, we consider that the real system parameters are 20% higher than those in Table I, and are used for evaluating the attack and the detection schemes. We consider the case of symmetric information availability, i.e., the operator and the attacker have access to the same inaccurate parameters. The parameters available are the ones shown in Table I, and are used for computing the UIO matrices, the residuals, for training DRLA s. There is thus 20% estimation error, which would not drastically increase the frequency deviations or UIO residuals without an attack, but is large enough to affect detectability.

Table III presents the attack impact and the AUC achieved by the detectors in this scenario. Surprisingly, even though the attacker uses the same inaccurate parameters as the operator, the attack impact is significantly increased for DRLA s compared to Table II, while remaining completely undetectable w.r.t. most detectors. Observe that the AUC for the UIO and ACE detectors are significantly smaller than 0.5 for some attacks. This means that DRLA learns to yield UIO residuals and ACE values that are on average smaller than the no-attack case.

Overall, our results indicate that DRLA s are powerful w.r.t. both the inflicted impact to the power grid, and the stealthiness against a wide range of detectors. However, the results also suggest potential methods to enhance the security of AGC, including (1) obtaining more accurate system models and information, (2) utilizing supervised ML detectors with rich training data, and (3) securing measurements from physical and network intrusions, by, e.g., utilizing redundant frequency measurements.

VI. Conclusion

In this paper we investigated the vulnerability of state-of-the-art AGC to attacks against power and frequency measurements. We formulated the problem of attacking an AGC system equipped with multiple fault and attack detection methods as a POMDP. We proposed an RL solution based on the proximal policy optimization algorithm to compute the attacked sensor measurements. Our results show the superiority of the proposed RL-based attacks compared to several baseline attacks in terms of stealthiness and attack impact, and show that sophisticated attacks could bypass existing detection schemes and could lead the grid frequency to critical trajectories. One direction for future work could be to analyze the practical feasibility of the proposed attack when considering weaker attack models, e.g., attackers without knowledge of the system parameters, or those manipulating measurements in only one area.

Appendix

A. *FDIA Detection using UIO and CUSUM*

In what follows we consider that the system operator is using a combination of the UIO detector and CUSUM (i.e., referred to as the *CUSUM(UIO) detector*). The detection metric for the CUSUM(UIO) detector is computed as [66]

$$\mathcal{S}_r[t] = \max\ (0,\ \mathcal{S}_r[t-1] + \|r[t]\|_2 - b_r), \tag{29}$$

where $\|r[t]\|_2$ is the ℓ_2-norm of the UIO residual at time t, b_r is the bias term chosen to be equal to the mean UIO residual in the normal (unattacked) case, and $\mathcal{S}_r[0] = 0$. An alarm is then raised by the detector if

$$\mathcal{S}_r[t] > \rho_c, \tag{30}$$

where ρ_c is a predefined detection threshold.

Using the same simulated data described in Section V-B, we evaluated the CUSUM(UIO) detector against the baseline FDIAs and our proposed DRLAs, and the results are shown in Figure 7 and Figure 8. The figures show the trade-off between the attack impact, and the maximum CUSUM detection metric ($\mathcal{S}_r$) during an episode. For the case when only power measurements are attacked, Figure 7 shows that using CUSUM can improve the separability of the attacked and the non-attacked measurements, compared to the UIO-detector which directly uses the raw residuals (c.f., Figure 3). To the contrary, when both power and frequency measurements are attacked, Figure 8 shows that the CUSUM(UIO) detector did not bring any performance improvement compared to the UIO detector (c.f., Figure 4). Observe that in the former case, the attacked UIO residuals were on average higher than the non-attacked ones. Using CUSUM in that case allows this difference to accumulate over time, and thus boosts the detection performance. On the other hand, the attacked UIO residuals were on average less than or equal to the non-attacked residuals in the latter case. Thus, using CUSUM makes little to no difference in the detection performance.

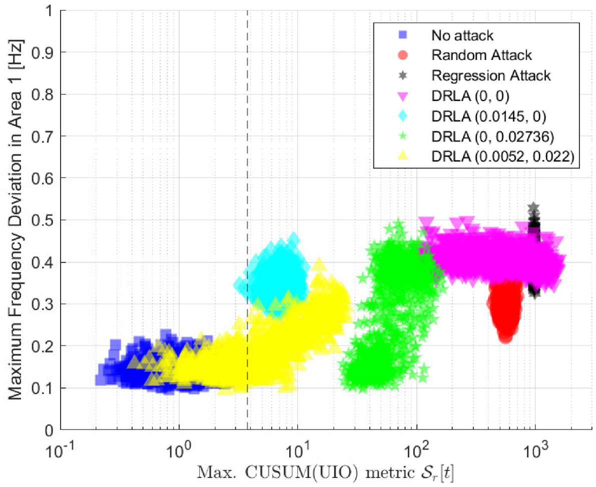

Fig. 7. Trade-off between attack impact and the CUSUM detection metric for DRLAs and baselines, when only power-flow measurements are attacked.

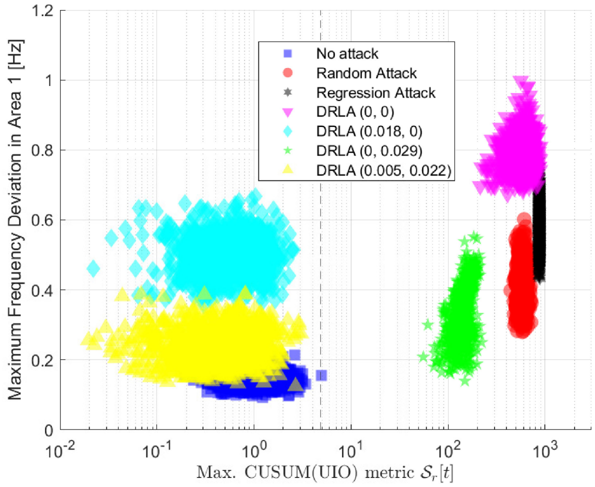

Fig. 8. Trade-off between attack impact and the CUSUM detection metric for DRLAs and baselines, when both power-flow and frequency measurements are attacked.

Note that the DRLAs in Figure 8 were capable of bypassing the CUSUM(UIO) detector despite the fact that they were trained to minimize $\|r[t]\|_2$ and not $\mathcal{S}_r[t]$. Training DRLAs that target $\mathcal{S}_r[t]$ should yield even stealthier attacks. Furthermore, one could also implement a *CUSUM(ACE) detector*, but our results suggest that such a detector would provide little improvement in detection performance, especially for the case when both power and frequency measurements are attacked.

References

[1] P. P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.

[2] H. Sadat, *Power System Analysis*. New York, NY, USA: McGraw-Hill, 2004.

[3] R. Tan et al., "Modeling and mitigating impact of false data injection attacks on automatic generation control," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1609–1624, Jul. 2017.

[4] S. Sundaram and C. N. Hadjicostis, "Delayed observers for linear systems with unknown inputs," *IEEE Trans. Autom. Control*, vol. 52, no. 2, pp. 334–339, Feb. 2007.

[5] H. Soh and Y. Demiris, "Evolving policies for multi-reward partially observable Markov decision processes (MR-POMDPs)," in *Proc. Ann. Conf. Genet. Evol. Comput.*, 2011, pp. 713–720.

[6] K. Rahimi, A. Parchure, V. Centeno, and R. Broadwater, "Effect of communication time-delay attacks on the performance of automatic generation control," in *Proc. North Amer. Power Symp. (NAPS)*, 2015, pp. 1–6.

[7] X. Lou et al., "Assessing and mitigating impact of time delay attack: Case studies for power grid controls," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 141–155, Jan. 2020.

[8] S. Sridhar and G. Manimaran, "Data integrity attacks and their impacts on SCADA control system," in *Proc. IEEE PES Gen. Meeting*, 2010, pp. 1–6.

[9] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 580–591, Mar. 2014.

[10] C. Chen, X. Zhang, M. Cui, K. Zhang, J. Zhao, and F. Li, "Stability assessment of secondary frequency control system with dynamic false data injection attacks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3224–3234, May 2022.

[11] C. Chen, M. Cui, X. Wang, K. Zhang, and S. Yin, "An investigation of coordinated attack on load frequency control," *IEEE Access*, vol. 6, pp. 30414–30423, 2018.

[12] W. Yan et al., "A stealthier false data injection attack against the power grid," in *Proc. IEEE Int. Conf. Commun. Control Comput. Technol. Smart Grids (SmartGridComm)*, 2021, pp. 108–114.

[13] M. Jafari, M. A. Rahman, and S. Paudyal, "Optimal false data injection attacks against power system frequency stability," *IEEE Trans. Smart Grid*, vol. 14, no. 2, pp. 1276–1288, Mar. 2023.

[14] S. D. Roy and S. Debbarma, "Detection and mitigation of cyber-attacks on AGC systems of low inertia power grid," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2023–2031, Jun. 2020.

[15] X. He, X. Liu, and P. Li, "Coordinated false data injection attacks in AGC system and its countermeasure," *IEEE Access*, vol. 8, pp. 194640–194651, 2020.

[16] F. Zhang and Q. Li, "Deep learning-based data forgery detection in automatic generation control," in *Proc. IEEE Conf. Commun. Netw. Security (CNS)*, 2017, pp. 400–404.

[17] C. Chen, K. Zhang, K. Yuan, L. Zhu, and M. Qian, "Novel detection scheme design considering cyber attacks on load frequency control," *IEEE Trans. Ind. Informat.*, vol. 14, no. 5, pp. 1932–1941, Sep. 2019.

[18] Z. Chen, J. Zhu, S. Li, Y. Liu, and T. Luo, "Detection of false data injection attacks on load frequency control system with renewable energy based on fuzzy logic and neural networks," *J. Mod. Power Syst. Clean Energy*, vol. 10, no. 6, pp. 1576–1587, 2022.

[19] A. Ameli, A. Hooshyar, E. F. El-Saadany, and A. M. Youssef, "Attack detection and identification for automatic generation control systems," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4760–4774, Sep. 2018.

[20] A. Ameli, A. Hooshyar, A. H. Yazdavar, E. F. El-Saadany, and A. Youssef, "Attack detection for load frequency control systems using stochastic unknown input estimators," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 10, pp. 2575–2590, Oct. 2018.

[21] 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, Sep. 2019.

[22] J. Ye and X. Yu, "Detection and estimation of false data injection attacks for load frequency control systems," *J. Mod. Power Syst. Clean Energy*, vol. 10, no. 4, pp. 861–870, 2022.

[23] K. Xiahou, Y. Liu, and Q. H. Wu, "Decentralized detection and mitigation of multiple false data injection attacks in multiarea power systems," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 3, no. 1, pp. 101–112, Jan. 2022.

[24] X. Chen, S. Hu, Y. Li, D. Yue, C. Dou, and L. Ding, "Co-estimation of state and FDI attacks and attack compensation control for multi-area load frequency control systems under FDI and DoS attacks," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2357–2368, May 2022.

[25] K. Pan, E. Rakhshani, and P. Palensky, "False data injection attacks on hybrid AC/HVDC interconnected systems with virtual inertia—Vulnerability, impact and detection," *IEEE Access*, vol. 8, pp. 141932–141945, 2020.

[26] A. D. Syrmakesis, H. H. Alhelou, and N. D. Hatziargyriou, "Novel SMO-based detection and isolation of false data injection attacks against frequency control systems," *IEEE Trans. Power Syst.*, early access, Feb. 3, 2023, doi: 10.1109/TPWRS.2023.3242015.

[27] M. Khalaf, A. Youssef, and E. El-Saadany, "Detection of false data injection in automatic generation control systems using Kalman filter," in *Proc. IEEE Elect. Power Energy Conf. (EPEC)*, 2017, pp. 1–6.

[28] A. S. L. V. Tummala and R. K. Inapakurthi, "A two-stage Kalman filter for cyber-attack detection in automatic generation control system," *J. Mod. Power Syst. Clean Energy*, vol. 10, no. 1, pp. 50–59, 2022.

[29] M. Khalaf, A. Youssef, and E. El-Saadany, "A particle filter-based approach for the detection of false data injection attacks on automatic generation control systems," in *Proc. IEEE Elect. Power Energy Conf. (EPEC)*, 2018, pp. 1–6.

[30] T. Huang, B. Satchidanandan, P. R. Kumar, and L. Xie, "An online detection framework for cyber attacks on automatic generation control," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6816–6827, Nov. 2018.

[31] S. D. Roy, S. Debbarma, and A. Iqbal, "A decentralized intrusion detection system for security of generation control," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18924–18933, Oct. 2022.

[32] Y. Li, R. Huang, and L. Ma, "False data injection attack and defense method on load frequency control," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2910–2919, Feb. 2021.

[33] A. S. Musleh, G. Chen, Z. Y. Dong, C. Wang, and S. Chen, "Attack detection in automatic generation control systems using LSTM-based stacked autoencoders," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 153–165, Jan. 2023.

[34] A. Abbaspour, A. Sargolzaei, P. Forouzannezhad, K. K. Yen, and A. I. Sarwat, "Resilient control design for load frequency control system under false data injection attacks," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 7951–7962, Sep. 2020.

[35] J. Wang, D. Wang, L. Su, J. H. Park, and H. Shen, "Dynamic event-triggered H_∞ load frequency control for multi-area power systems subject to hybrid Cyber attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 12, pp. 7787–7798, Dec. 2022.

[36] C. Chen, Y. Chen, J. Zhao, K. Zhang, M. Ni, and B. Ren, "Data-driven resilient automatic generation control against false data injection attacks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 8092–8101, Dec. 2021.

[37] 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.

[38] Z. Zhang, J. Hu, J. Lu, J. Cao, and F. E. Alsaadi, "Preventing false data injection attacks in LFC system via the attack-detection evolutionary game model and KF algorithm," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 6, pp. 4349–4362, Nov./Dec. 2022.

[39] T. Imthias Ahamed, P. Nagendra Rao, and P. Sastry, "A reinforcement learning approach to automatic generation control," *Elect. Power Syst. Res.*, vol. 63, no. 1, pp. 9–26, 2002.

[40] C. Watkins and P. Dayan, "Technical note: *Q*-learning," *Mach. Learn.*, vol. 8, pp. 279–292, May 1992.

[41] L. Xi, L. Yu, Y. Xu, S. Wang, and X. Chen, "A novel multi-agent DDQN-AD method-based distributed strategy for automatic generation control of integrated energy systems," *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2417–2426, Oct. 2020.

[42] H. v. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double *Q*-learning," in *Proc. AAAI*, 2016, pp. 2094–2100.

[43] S. Wang et al., "A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4644–4654, Nov. 2020.

[44] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor–critic for mixed cooperative-competitive environments," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 6382–6393.

[45] D. Chen et al., "PowerNet: Multi-agent deep reinforcement learning for scalable powergrid control," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1007–1017, Mar. 2022.

[46] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2158–2169, Mar. 2019.

[47] I. D. Landau, G. Zito, *Digital Control Systems: Design, Identification and Implementation*. London, U.K.: Springer, 2006.

[48] M. Sain and J. Massey, "Invertibility of linear time-invariant dynamical systems," *IEEE Trans. Autom. Control*, vol. AC-14, no. 2, pp. 141–149, Apr. 1969.

[49] I. N. Fovino, A. Carcano, T. De Lacheze Murel, A. Trombetta, and M. Masera, "Modbus/DNP3 state-based intrusion detection system," in *Proc. IEEE Int. Conf. Adv. Inf. Netw. Appl.*, 2010, pp. 729–736.

[50] C. Brunner, "IEC 61850 for power system communication," in *Proc. IEEE/PES Transm. Distrib. Conf. Expo.*, 2008, pp. 1–6.

[51] S. M. S. Hussain, T. S. Ustun, and A. Kalam, “A review of IEC 62351 security mechanisms for IEC 61850 message exchanges,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5643–5654, Sep. 2020.

[52] E. Shereen, M. Delcourt, S. Barreto, G. Dán, J.-Y. Le Boudec, and M. Paolone, “Feasibility of time-synchronization attacks against PMU-based state estimation,” *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3412–3427, Jun. 2020.

[53] D. Kushner, “The real story of Stuxnet,” *IEEE Spectr.*, vol. 50, no. 3, pp. 48–53, May 2013.

[54] Y. Liu, P. Ning, and M. K. Reiter, “False data injection attacks against state estimation in electric power grids,” in *Proc. ACM CCS*, 2009, pp. 21–32.

[55] G. Dán and H. Sandberg, “Stealth attacks and protection schemes for state estimators in power systems,” in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2010, pp. 214–219.

[56] S. S. Stanković, X.-B. Chen, M. R. Mataušek, and D. D. Šiljak, “Decentralized automatic generation control based on a stochastic inclusion principle,” in *Proc. IFAC/IFORS/IMACS/IFIP Symp. Large Scale Syst. Theory Appl.*, 1998, pp. 241–248.

[57] V. Mnih et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[58] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” 2017, *arXiv:1707.06347*.

[59] T. Lillicrap et al., “Continuous control with deep reinforcement learning,” Sep. 2015. [Online]. Available: https://arxiv.org/abs/1509.02971

[60] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor–critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *Proc. ICML*, 2018, pp. 1–6.

[61] S. Kullback, *Information Theory and Statistics*. New York, NY, USA: Wiley, 1959.

[62] E. Liang et al., “RLlib: Abstractions for distributed reinforcement learning,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3053–3062.

[63] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, “High-dimensional continuous control using generalized advantage estimation,” in *Proc. Int. Conf. Learn. Rep. (ICLR)*, 2016, p. 6.

[64] M. Abadi et al., “TensorFlow: A system for large-scale machine learning,” in *Proc. {USENIX} Symp. Oper. Syst. Design Implement. ({OSDI})*, 2016, pp. 265–283.

[65] S. Ruder, “An overview of gradient descent optimization algorithms,” 2016, *arxiv.abs/1609.04747*.

[66] C. Murguia and J. Ruths, “CUSUM and chi-squared attack detection of compromised sensors,” in *Proc. IEEE Conf. Control Appl. (CCA)*, 2016, pp. 474–480.

[67] A. Nisioti, A. Mylonas, P. D. Yoo, and V. Katos, “From intrusion detection to attacker attribution: A comprehensive survey of unsupervised methods,” *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3369–3388, 4th Quart., 2018.

[68] M. Z. Alom and T. M. Taha, “Network intrusion detection for cyber security using unsupervised deep learning approaches,” in *Proc. IEEE Nat. Aerosp. Electron. Conf. (NAECON)*, 2017, pp. 63–69.

[69] J. Zhang and M. Zulkernine, “Anomaly based network intrusion detection with unsupervised outlier detection,” in *Proc. IEEE Int. Conf. Commun.*, vol. 5, 2006, pp. 2388–2393.

[70] J. Zhang, L. Pan, Q.-L. Han, C. Chen, S. Wen, and Y. Xiang, “Deep learning based attack detection for cyber-physical system cybersecurity: A survey,” *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 3, pp. 377–391, Mar. 2022.

Ezzeldin Shereen (Member, IEEE) received the B.Sc. and M.Sc. degrees in networking engineering from the German University in Cairo, Egypt, in 2014 and 2015, respectively, and the Ph.D. degree from the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, in 2021, where he is currently working as a Postdoctoral Researcher. His research interests include cyberphysical systems security with focus on power systems, robustness of machine learning and reinforcement learning, and performance evaluation of wireless networks.

Kiarash Kazari (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering (communication systems branch) from the Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree with the KTH Royal Institute of Technology, Stockholm, Sweden, conducting research on the security aspects of multi-agent learning systems. His research interests include communication networks, multiagent learning systems, and reinforcement learning.

György Dán (Senior Member, IEEE) received the M.Sc. degree in computer engineering from the Budapest University of Technology and Economics, Budapest, Hungary, in 1999, the M.Sc. degree in business administration from the Corvinus University of Budapest, Budapest, in 2003, and the Ph.D. degree in telecommunications from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2006. From 1999 to 2001, he was a Consultant in the field of Access Networks, Streaming Media, and Videoconferencing with BCN Ltd., Budapest. He was a Visiting Researcher with the Swedish Institute of Computer Science, Stockholm, in 2008, a Fulbright Research Scholar with the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2012 and 2013, and an Invited Professor with the Swiss Federal Institute of Technology of Lausanne, Lausanne, Switzerland, in 2014 and 2015. He is currently a Professor with the KTH Royal Institute of Technology. His current research interests include the design and analysis of content management and computing systems, game theoretical models of networked systems, and cyber–physical system security and resilience. He has been an Area Editor of *Computer Communications* from 2014 to 2021 and the IEEE TRANSACTIONS ON MOBILE COMPUTING from 2019 to 2023.