Adversarial Robustness of Multi-agent Reinforcement Learning Secondary Control of Islanded Inverter-based AC Microgrids

Ezzeldin Shereen, Kiarash Kazari and György Dán
Division of Network and Systems Engineering, School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology, Stockholm, Sweden
E-mail:{eshereen, kkazari, gyuri}@kth.se

***Abstract*—Secondary control of voltage magnitude and frequency is essential to the stable and secure operation of microgrids (MGs). Recent years have witnessed an increasing interest in developing secondary controllers based on multi-agent reinforcement learning (MARL), in order to replace existing model-based controllers. Nonetheless, unlike the vulnerabilities of model-based controllers, the vulnerability of MARL-based MG secondary controllers has so far not been addressed. In this paper, we investigate the vulnerability of MARL controllers to false data injection attacks (FDIAs). Based on a formulation of MG secondary control as a partially observable stochastic game (POSG), we propose to formulate the problem of computing FDIAs as a partially observable Markov decision process (POMDP), and we use state-of-the-art RL algorithms for solving the resulting problem. Based on extensive simulations of a MG with 4 distributed generators (DGs), our results show that MARL-based secondary controllers are more resilient to FDIAs compared to state of the art model-based controllers, both in terms of attack impact and in terms of the effort needed for computing impactful attacks. Our results can serve as additional arguments for employing MARL in future MG control.**

I. Introduction

Microgrids (MG) are emerging as a means of integrating distributed generators (DG) and energy storage systems (ESSs) into conventional power systems. A microgrid could include conventional rotating machinery DGs, e.g., synchronous power generators, but could also include DC sources such as solar panels and fuel cells. In addition to working in grid-connected mode where the power generation as well as voltage and frequency control are supported by the main power grid, microgrids can work in islanded mode either during planned maintenance times, or during unexpected faults and outages. While islanded operation can increase reliability, maintaining voltage and frequency stability becomes challenging, and has to be achieved relying on local control loops only [1].

Voltage and frequency control in microgrids are typically hierarchical, consisting of primary, secondary, and tertiary control. Primary control uses the droop-control method widely employed with synchronous generators to create damping in the system. The secondary control ensures that voltage and frequency deviations caused by primary control stay within acceptable limits. Tertiary control, utilized in grid-connected mode, controls the power flow between the microgrid and the main grid. Primary control is typically implemented locally, as an integrated control loop in each DG. On the contrary, secondary voltage and frequency control have been conventionally implemented in a centralized manner at a microgrid centralized controller (MGCC) [2]. Nonetheless, the MGCC introduces a single point of failure in the network, and there is thus a recent interest in distributed secondary controllers that rely on communications between neighbouring DGs [3, 4].

Early works on distributed secondary control of inverter-based microgrids followed a model-based approach, and there is a rich literature on their vulnerability to cyber-attacks on MG sensors, such as false data injection attacks (FDIAs) [5]. Similarily, effective attack detectors and fault-tolerant control algorithms have been proposed to mitigate known vulnerabilities [6, 7]. The main detriment of model-based controllers is that they rely on complex models, which are often inaccurate, and hence they are sensitive to system topology and parameter changes [8], which has led to an increasing interest for multi-agent reinforcement learning (MARL) based controllers. Unfortunately, even though it is known that MARL is vulnerable to adversarial attacks against one or many of the participating agents [9], the vulnerability of MARL-based secondary controllers has so far been not been investigated.

In this paper, we address this important issue and we make the following main contributions:

1) We conduct the first study to evaluate the vulnerability of MARL-based secondary controllers in inverter-based microgrids to FDIAs.
2) We formulate the problem of computing worst-case FDIAs against MG secondary control as a partially observable Markov decision process (POMDP).
3) We carry out extensive simulations to assess the impact of FDIAs on MARL-based secondary controllers, and compare the results to those obtained on traditional model-based controllers. Our results represent a first step towards making MARL-based MGs trustworthy.

The rest of this paper is organized as follows. Section II discusses the previous work on distributed MG secondary controllers and their vulnerability. The considered model of an inverter-based MG as well as the attacker model are presented in Section III. Section IV formulates the problem of attacking the controller as a partially observable Markov decision process. Section V shows the effect of different attacks on MARL controllers. Finally, Section VI concludes the paper.

II. Related Work

Vulnerability of MGs: A large body of works has investigated the security of MGs and their vulnerability to attacks, such as denial-of-service (DoS) attacks, time delay attacks (TDAs), and false data injection attacks (FDIAs) [10]. For example, [11] considers DoS attacks against the most critical DG in an islanded MG. Moreover, the effect of time delays due to communication delays (or delay attacks) on the accuracy on MG secondary control is investigated in [12]. However, the most commonly studied attack against inverter-based MGs is the FDIA. For example, authors in [5] show that FDIAs can have disastrous impact on frequency and voltage stability in MGs.

Several recent works have considered detecting FDIAs against MGs using model-based detectors, such as the unknown input observer (UIO). For example, [6] performed dynamic state estimation of an islanded MG using a UIO in order to detect and identify FDIAs against frequency measurements. Authors in [13] proposed a subspace-based detection technique (similar to a UIO) to detect FDIAs. Contrary to model-based approaches, few research works considered using

data-driven and machine learning based solutions to detect attacks against inverter-based MGs. Authors in [14] proposed an entropy-based attack detection approach for distributed secondary control of MGs, utilizing the Kullback-Leibler (KL) divergence between the distribution of control variables before and after the attack. Authors in [15] developed multiple machine learning models to detect a special type of FDIAs, called measurement-as-reference (MaR) attacks.

Going beyond attack detection, several works have considered fault-tolerant and resilient control for inverter-based MGs in order to alleviate the effect of attacks. These works typically combine attack detection and mitigation. An example is [16], which considers the problem of resilient secondary control under communication faults (including cyber-attacks) and designs a distributed observer for each DG that estimates the reference voltage and frequency values, and restores them to the reference values. In [7], every DG keeps a trust factor (i.e., a score) for each of its neighbouring DGs, which helps mitigate the effect of the attack. Similarly, [17] proposes a systematic approach to discard information from non-cooperative DGs. However, the above two approaches require certain connectivity criterion on the communication network graph. Moreover, most of the above works on detection and resilient control are specifically developed for model-based distributed secondary controllers, and hence might not be applicable to MGs employing MARL-based controllers.

MARL-based secondary controllers: Several recent works considered using reinforcement learning (RL) and multi-agent RL (MARL) for MG secondary control. Recently, [18] used the multi-agent deep deterministic policy gradient (MADDPG) algorithm for secondary frequency control of islanded MGs. Authors in [8] proposed another actor-critic MARL algorithm for voltage control in MGs, and compared their approach to multiple centralized and distributed RL algorithms. Despite the recent interest for MARL-based secondary control of MGs, to the best of our knowledge, this is the first work to assess their vulnerability to FDIAs.

III. System Model

A. Model of an Inverter-based Microgrid

We consider a microgrid with a set $\mathcal{G}$ of DGs, where $N = |\mathcal{G}|$ is the number of DGs. Furthermore, the microgrid has M lines, B buses L loads. A microgrid is typically modelled by considering its three interacting parts: the DGs, the network, and the loads [19]. The DG model includes the inverter, an LC filter, an output connector, as well as power, voltage, and current control loops, as shown in Figure 1. The network model specifies the connectivity of different DGs, and the load model contains the loads attached to each bus in the network. In what follows we provide the model for each component.

1) DG Model: Figure 1 shows a block diagram of the DG model using a voltage source inverter. The power part includes a three-leg inverter, an output LC filter and a coupling conductor. For this part, the inverter is typically assumed to be a perfect power source, and the DC bus dynamics as well as the inverter switching process can be neglected [19]. The control part includes a power control loop, which controls the voltage magnitude and frequency according to the droop mechanism mimicking the behaviour of a synchronous generator. It also includes voltage and current controllers, which are designed to reject high frequency disturbances and provide damping for the output LC filter.

The dynamics of each DG is typically given in its own direct-quadrature (d-q) reference frame. The angle δ_i of the reference frame of DG i w.r.t. the common reference frame follows the differential equation

$$\dot{\delta}_i = \omega_i - \omega_{com}, \tag{1}$$

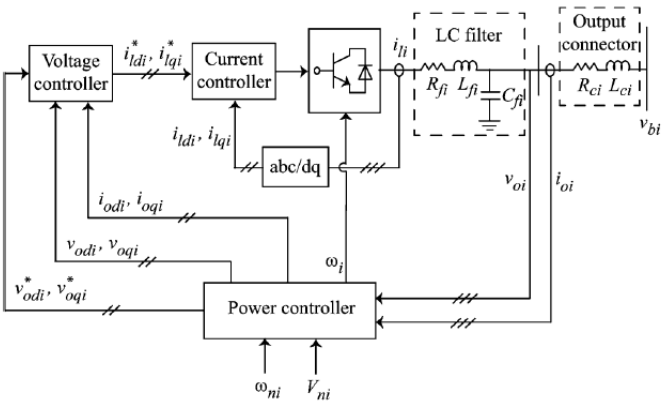

Figure 1: Block diagram of the control of a voltage-source inverter [20].

where ω_i is the angular frequency of the ith DG reference frame, and ω_{com} is the angular frequency of the common reference frame, which is chosen to be one of the DGs (in this paper, $\omega_{com} = \omega_1$). In the following we provide the dynamics for each of the separate components of the DG model.

Power controller: A traditional synchronous generator will react to an increase in the load by reducing its rotation frequency. For inverter-based microgrids, the power controller aims to mimic this behavior by defining a relation between the frequency and the active power, and between the voltage magnitude and the reactive power. First, the active and reactive power follow the dynamics

$$\dot{P} = -\omega_c P + \omega_c (v_{odi} i_{odi} + v_{oqi} i_{oqi}), \tag{2}$$

$$\dot{Q} = -\omega_c Q + \omega_c (v_{oqi} i_{odi} - v_{odi} i_{oqi}), \tag{3}$$

where ω_c is the cut-off frequency of two low-pass filters used to obtain the real and reactive powers corresponding to the fundamental components, and $v_{odi}, i_{odi}, v_{oqi}, i_{oqi}$ are the d and q components of the output voltage and currents shown in Figure 1. Based on the values of P and Q, the power controller computes the reference voltage magnitude v^*_{oi} supplied to the voltage controller, as well as the frequency supplied to the inverter according to the droop mechanism

$$\omega_i = \omega_{ni} - m_{Pi} P_i, \tag{4}$$

$$v^*_{odi} = V_{ni} - n_{Qi} Q_i, \tag{5}$$

$$v^*_{oqi} = 0, \tag{6}$$

where m_{Pi} and n_{Qi} are droop coefficients, and ω_{ni} and V_{ni} are the reference voltage frequency and amplitude values supplied by the secondary controller. If secondary control is not implemented, then $\omega_{ni} \leftarrow \omega_{ref}$ and $V_{ni} \leftarrow V_{ref}$, where the reference values are the nominal frequency and voltage values.

Voltage Controller: The voltage controller is responsible for supplying the reference current i^*_{li} to the current controller and is implemented as a PI-controller,

$$i^*_{ldi} = F_i i_{odi} - \omega_b C_{fi} v_{oqi} + K_{pvi}(v^*_{odi} - v_{odi}) + K_{ivi}\phi_{di}, \tag{7}$$

$$i^*_{lqi} = F_i i_{oqi} + \omega_b C_{fi} v_{odi} + K_{pvi}(v^*_{oqi} - v_{oqi}) + K_{ivi}\phi_{qi}, \tag{8}$$

where F_i is a constant, ω_b is the nonimal angular frequency (i.e., $2\pi * 60$ in the US.), K_{pvi} and K_{ivi} are the proportional and integral gains of the voltage controller, and ϕ_{di} and ϕ_{qi} are the integral error variables of the PI-controller, following the dynamics

$$\dot{\phi}_{di} = v^*_{odi} - v_{odi}, \qquad \dot{\phi}_{qi} = v^*_{oqi} - v_{oqi}. \tag{9}$$

Current Controller: The current controller is responsible for supplying the reference input voltage v_{ii}^* to the inverter and is implemented as a PI-controller,

$$v_{idi}^* = -\omega_b L_{fi} i_{lqi} + K_{pci}(i_{ldi}^* - i_{ldi}) + K_{ici}\gamma_{di}, \quad (10)$$

$$v_{iqi}^* = \omega_b L_{fi} i_{ldi} + K_{pci}(i_{lqi}^* - i_{lqi}) + K_{ici}\gamma_{qi}, \quad (11)$$

where K_{pci} and K_{ici} are the proportional and integral gains of the current controller, and γ_{di} and γ_{qi} are the integral error variables of the PI-controller, following the dynamics

$$\dot{\gamma}_{di} = i_{ldi}^* - i_{ldi}, \qquad \dot{\gamma}_{qi} = i_{lqi}^* - i_{lqi}. \quad (12)$$

Output LC Filter and Output Connector: Assuming an ideal inverter that produces the demanded voltage (i.e., $v_i = v_i^*$), and that DG i is connected to bus j, the dynamics of the output LC filter and output connector are given by

$$\dot{i}_{ldi} = \frac{-R_{fi}}{L_{fi}} i_{ldi} + \omega_i i_{lqi} + \frac{1}{L_{fi}}(v_{idi} - v_{odi}), \quad (13)$$

$$\dot{i}_{lqi} = \frac{-R_{fi}}{L_{fi}} i_{lqi} - \omega_i i_{ldi} + \frac{1}{L_{fi}}(v_{iqi} - v_{oqi}), \quad (14)$$

$$\dot{v}_{odi} = \omega_i v_{oqi} + \frac{1}{C_{fi}}(i_{ldi} - i_{odi}), \quad (15)$$

$$\dot{v}_{oqi} = -\omega_i v_{odi} + \frac{1}{C_{fi}}(i_{lqi} - i_{oqi}), \quad (16)$$

$$\dot{i}_{odi} = \frac{-R_{ci}}{L_{ci}} i_{odi} + \omega_i i_{oqi} + \frac{1}{L_{ci}}(v_{odi} - v_{bdj}), \quad (17)$$

$$\dot{i}_{oqi} = \frac{-R_{ci}}{L_{ci}} i_{oqi} - \omega_i i_{odi} + \frac{1}{L_{ci}}(v_{oqi} - v_{bqj}), \quad (18)$$

where (R_{fi}, L_{fi}, C_{fi}) are the resistance, inductance, and capacitance of the LC filter, (R_{ci}, L_{ci}) are the resistance and inductance of the output connector, and v_{bi} is the bus voltage at bus i.

2) Network Model: The network model captures the interaction between variables in different DGs. An example microgrid with $N = 4$ DGs, $M = 3$ lines, $B = 4$ buses, and $L = 2$ loads is shown in Figure 2. In general, consider that line i connects buses j and k. The dynamics of the current $i_{line,i}$ flowing across the line are given as

$$\dot{i}_{lineDi} = \frac{-R_{li}}{L_{li}} i_{lineDi} + \omega_i i_{lineQi} + \frac{1}{L_{l1}}(v_{bDj} - v_{bDk}), \quad (19)$$

$$\dot{i}_{lineQi} = \frac{-R_{li}}{L_{li}} i_{lineQi} - \omega_i i_{lineDi} + \frac{1}{L_{l1}}(v_{bQj} - v_{bQk}), \quad (20)$$

where (R_{li}, L_{li}) are the resistance and inductance of line i, and $(i_{lineDi}, i_{lineDi}, v_{bDj}, v_{bQj})$ are variables converted to the common reference frame. To convert variables from each DG's reference frame to the common reference frame, Park's transformation [19] is used,

$$x_{DQi} = T_i x_{dqi}, \quad (21)$$

where

$$T_i = \begin{bmatrix} \cos\delta_i & -\sin\delta_i \\ \sin\delta_i & \cos\delta_i \end{bmatrix}, x_{DQi} = [x_{Di}, x_{Qi}]^T, x_{dqi} = [x_{di}, x_{qi}]^T. \quad (22)$$

3) Load Model: Considering that load i is attached to bus j, the dynamics of the current drawn by the load is given by

$$\dot{i}_{loadDi} = \frac{-R_{Li}}{L_{Li}} i_{loadDi} + \omega_i i_{loadQi} + \frac{1}{L_{L1}} v_{bDj}, \quad (23)$$

$$\dot{i}_{loadQi} = \frac{-R_{Li}}{L_{Li}} i_{loadQi} - \omega_i i_{loadDi} + \frac{1}{L_{L1}} v_{bQj}, \quad (24)$$

where (R_{Li}, L_{Li}) are the resistance and inductance of load i. Finally, the voltages at all buses $v_{bDQ} \in \mathbb{R}^{2B}$ could be computed using the matrix equation

$$v_{bDQ} = R_N(M_{inv} i_{oDQ} + M_{load} i_{loadDQ} + M_{net} i_{lineDQ}), \quad (25)$$

where R_N is a virtual resistance assumed between each bus and the ground and is chosen to be sufficiently large to have minimum impact on the system stability [19], $i_{oDQ} \in \mathbb{R}^{2N}$, $i_{loadDQ} \in \mathbb{R}^{2L}$, and $i_{lineDQ}) \in \mathbb{R}^{2M}$. Furthermore, $M_{inv} \in \{0,1\}^{2B \times 2N}$ is a mapping matrix which maps the DGs to the network buses. $M_{load} \in \{-1,0\}^{2B \times 2L}$ is a mapping matrix which maps the load locations (-1) to the network buses. Finally, $M_{net} \in \{-1,0,1\}^{2B \times 2M}$ is a mapping matrix which maps the lines to the network buses. M_{net} includes a (+1) if the current is entering the node, and (-1) if the current is leaving the bus.

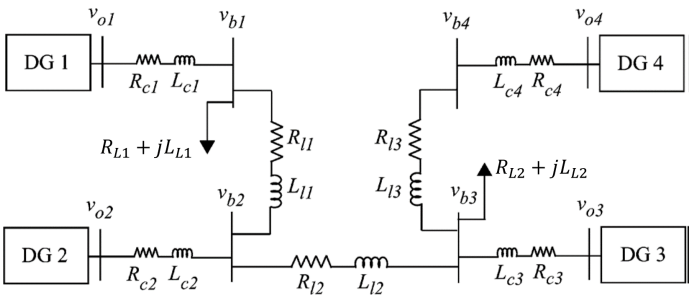

Figure 2: Single-line diagram of a Microgrid with four DGs [20].

B. Secondary control of Inverter-based Microgrids

The main objective of secondary control is to drive the voltage and frequency deviations caused by primary control to their nominal values. In what follows we discuss three approaches for implementing secondary control in inverter-based microgrids; centralized controllers, decentralized model-based controllers,and decentralized multi-agent reinforcement learning (MARL)-based controllers.

1) Microgrid Centralized Controller (MGCC): The MGCC focuses on regulating the voltage and frequency of the so-called critical bus (i.e., the bus hosting the most critical loads in the microgrid), and uses a PI-controller to compute the voltage and frequency compensation terms as

$$\delta\omega = K_{p\omega}(\omega_{ref} - \omega_{cr}) + K_{i\omega}\beta_\omega, \quad (26)$$

$$\delta V = K_{pV}(V_{ref} - v_{ocr}) + K_{iV}\beta_V, \quad (27)$$

where $(K_{p\omega}, K_{i\omega}, K_{pV}, K_{iV})$ are the MGCC PI-controller gains, (ω_{cr}, v_{ocr}) are the inverter output voltage frequency and magnitude measurements at the critical bus, and (β_ω, β_V) are the respective integral error terms following the dynamics

$$\dot{\beta}_\omega = (\omega_{ref} - \omega_{cr}), \qquad \dot{\beta}_V = (V_{ref} - v_{ocr}). \quad (28)$$

The compensation terms $(\delta\omega, \delta V)$ are then sent to all DGs in the microgrid to adjust their frequency and voltage.

2) Distributed Model-based Controller: The most commonly used distributed secondary frequency and voltage controller is the model predictive controller (MPC) [1, 4]. The controller is based on voltage magnitude and frequency measurements communicated between neighbouring DGs. The secondary control references values are then computed as

$$\dot{\omega}_{ni} = -c_\omega \left(g_i(\omega_i - \omega_{ref}) + \sum_{j \in N_i} a_{ij}(\omega_i + m_{Pi}P_i - \omega_j - m_{Pj}P_j) \right), \quad (29)$$

$$\dot{V}_{ni} = -c_v \left(g_i(v_{odi} - v_{ref}) + \sum_{j \in N_i} a_{ij}(v_{odi} + n_{Qi}Q_i - v_{odj} - n_{Qj}Q_j) \right), \quad (30)$$

where c_ω and c_v are control gains, g_i is the so-called pinning gain of DG i, N_i is the set of neighbours of DG i, and $A = [a_{ij}] \in \{0,1\}^{N \times N}$ is the adjacency matrix of the communication graph of the MG.

3) MARL Secondary Controller: We consider a MARL controller akin to those considered in recent literature [8, 18], where each DG acts as an independent RL agent. The problem of learning a MARL secondary controller can be formulated as a partially-observable stochastic game (POSG) given by the tuple $M_c \triangleq (N, \mathcal{S}, \mathcal{A}^i, \mathcal{P}, R^i, \mathcal{O}^i, \gamma)$, where:

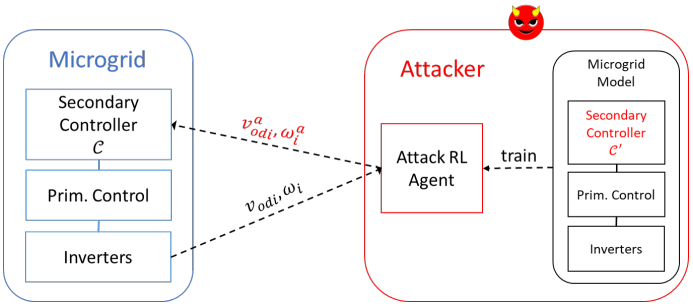

Figure 3: The considered attack model against inverter-based microgrids.

- N is the number of agents (i.e., DGs).
- $\mathcal{S}$ is the state space of the system, and $s_t \in \mathcal{S}$ is the state at time step t. For an inverter-based microgrid, the system state includes the variables $(\delta_i, P_i, Q_i, \phi_{di}, \phi_{qi}, \gamma_{di}, \gamma_{qi}, i_{ldi}, i_{lqi}, i_{odi}, i_{oqi}, v_{odi}, v_{oqi})$ for all DGs, the voltages (v_{bdi}, v_{bqi}) at all buses, the current (i_{lineDi}, i_{lineQi}) flowing in all lines, the current (i_{loadDi}, i_{loadDi}) flowing in all loads, as well as the load demand (R_{Li}, L_{Li}).
- $\mathcal{A}^i$ is the set of each agent's possible actions, and $a^i_t \in \mathcal{A}^i$ denotes the action of agent i at time step t. We denote the joint action of all agents by $\boldsymbol{a}_t = [a^1_t, ..., a^N_t] \in \boldsymbol{\mathcal{A}}$, where $\boldsymbol{\mathcal{A}} = \mathcal{A}^1 \times ... \times \mathcal{A}^N$. In this paper, we follow [8] and consider only the problem of voltage magnitude control, and assume that the MPC controller [4] in (29) is used to control the frequency. Therefore, the action of each agent includes the reference values $(a^i_t = V_{ni})$ supplied to the power controller in (5).
- $\mathcal{P}(s_{t+1} | s_t, a_t)$ is the conditional transition probability between states, determined by the dynamics explained in Section III-A. The dynamics is not known to the agents.
- R^i is the reward function of each agent. We adopt the reward of [8] which is designed to encourage the agents to keep the output voltage of each DG close to its nominal values as follows

$$R^i = \begin{cases} 0.05 - |1 - v^{pu}_{oi}|, & v^{pu}_{oi} \in [0.95, 1.05], \\ -|1 - v^{pu}_{oi}|, & v^{pu}_{oi} \in [0.8, 0.95] \cup [1.05, 1.2], \\ -20, & \text{otherwise}, \end{cases} \tag{31}$$

  where v^{pu}_{oi} is the inverter output voltage in p.u. given as

$$v^{pu}_{oi} = \frac{\sqrt{(v_{odi})^2 + (v_{oqi})^2}}{V_{ref}}. \tag{32}$$

- $\mathcal{O}^i$ denotes the observation space of each agent, and $o^i_t \in \mathcal{O}^i$ is the observation made by agent i at time t. Although [8] used the full state of the DG as the observation of each agent (i.e., $o^i_t = (\delta_i, P_i, Q_i, i_{odi}, i_{oqi}, v_{bdi}, v_{bqi})$), we adopted a more realistic observation inspired by [21]. Our considered observation only includes the voltage magnitude and frequency measurements (v_{odi}, ω_i) at both time t and $t-1$ to account for the derivative. Thus, we consider that $o^i_t = (v_{odi}[t], v_{odi}[t-1], \omega_i[t], \omega_i[t-1])$.
- $\gamma \in [0,1)$ is the discount factor.

A policy $\pi^i : \mathcal{O}^i \to \mathcal{A}^i$ for solving M_c can be learned, e.g., using reinforcement learning [8, 18].

C. Attack Model

The considered attack model is shown in Figure 3. We consider an attacker that has knowledge of the topology of the attacked microgrid as well as the underlying communication network. Moreover, the attacker knows the parameters (R_{li}, L_{li}) of all lines, the parameters $(R_{fi}, L_{fi}, C_{fi}, R_{ci}, L_{ci})$ of all LC filters and output connectors, the droop parameters (m_{Pi}, n_{Qi}), the voltage and current controller gains $(K_{pvi}, K_{ivi}, K_{pci}, K_{ici})$, as well as the nominal voltage frequency and magnitude values (ω_{ref}, V_{ref}). Based on the attacker´s knowledge about the secondary controller $\mathcal{C}$ used by the operator, we distinguish between two attack models: we refer to an attacker that knows the controller used by the operator as a *strong attacker*, while we refer to an attacker that does not know the controller as a *weak attacker*. We furthermore assume that the attacker can eavesdrop on the voltage magnitude and frequency measurements taken by a subset $\mathcal{G}^e \subseteq \mathcal{G}$ of the generators.

The attacker can manipulate the measurements taken by a subset $\mathcal{G}^a \subseteq \mathcal{G}^e$ of the generators according to

$$\omega^a_i = \omega_i + a^\omega_i, \qquad \forall i \in \mathcal{G}^a, \tag{33}$$

$$v^a_{odi} = v_{odi} + a^v_i, \qquad \forall i \in \mathcal{G}^a, \tag{34}$$

where (ω^a_i, v^a_{odi}) are the attacked measurement values, and (a^ω_i, a^v_i) are the perturbations. We assume that the perturbations are constrained,

$$|a^\omega_i| \le a^\omega_{max}, \qquad |a^v_i| \le a^v_{max}. \tag{35}$$

The manipulated measurements will be used by the secondary controller possibly leading to the computation of erroneous reference values (ω_{ni}, V_{ni}), which could potentially cause unstable voltage and frequency trajectories.

Given the above knowledge and attack capabilities, the attacker could implement a FDIA against the microgrid as follows:

1) Develop a dynamical model of the microgrid based on the known system topology and parameters.
2) Develop a model-based or RL secondary controller C' for the microgrid.
3) Use the dynamical model and the controller C' within a reinforcement learning framework to learn optimal FDIAs against the microgrid.
4) Apply the learned attacks against the real microgrid.

Although the above assumes a powerful attacker, recent attacks on smart grids have shown that attackers are capable of obtaining critical information and launching crafted and sophisticated attacks [22]. For example, the system topology and parameters could be obtained through insiders, by employing reconnaissance activities, or by exploiting publicly available data [23]. Eavesdropping and manipulating measurements could be possible due to the use of public communication networks and the lack of authentication mechanisms in smart grid communications.

IV. RL-based Attacks on Microgrid Secondary Control

To assess the resilience of MARL-based MG secondary controllers, in what follows we propose to formulate the problem of computing worst-case attacks as a partially observable Markov decision process (POMDP), and propose to use reinforcement learning for obtaining a close to optimal attack policy. Our proposed POMDP formulation of the attacker´s problem is given by the tuple $M_a \triangleq (\bar{\mathcal{S}}, \bar{\mathcal{A}}, \bar{\mathcal{P}}, \bar{R}, \bar{\mathcal{O}}, \bar{\gamma})$, where

- $\bar{\mathcal{S}}$ is the state space of the system, and $\bar{s}_t \in \bar{\mathcal{S}}$ is the state at time step t. The state here is the same as that in the operator's problem, in addition to the secondary control references (ω_{ni}, V_{ni}).
- $\bar{\mathcal{A}}$ is the set of possible actions for the attacker, and $\bar{a}_t \in \bar{\mathcal{A}}$ is the action of the attacker at time step t. The attacker's action consists of the measurement perturbations $(a^\omega_i, a^v_i), \forall i \in \mathcal{G}^a$, as specified in (33) and (34), respectively.
- $\bar{\mathcal{P}}(\bar{s}_{t+1} | \bar{s}_t, \bar{a}_t)$ is the conditional state transition probability.
- $\bar{R}$ is the reward function of the attacker. We propose to use the reward function $\bar{R} = -\sum_{i=1}^N R^i$. That is, the attacker is attempting to maximize the total voltage deviation of all DGs.
- $\bar{\mathcal{O}}$ denotes the observation space of the attacker, and $\bar{o}_t \in \bar{\mathcal{O}}$ is the attacker's observation at time t. The attacker can observe the observations o^i_t of DGs $i \in \mathcal{G}^e$. In addition, the attacker can

observe both the attacked and unattacked measurements. In other words, $\bar{o}_t = \{(v_{odi}[t], v_{odi}[t-1], \omega_i[t], \omega_i[t-1], v^a_{odi}[t], v^a_{odi}[t-1], \omega^a_i[t], \omega^a_i[t-1]) : \forall i \in \mathcal{G}^e\}$.

- $\bar{\gamma} \in [0,1)$ is a discount factor.

The above POMDP can be solved using a single-agent RL algorithm with continuous state and action space, such as the proximal policy optimization (PPO) algorithm [24].

V. Numerical Results

In what follows we evaluate the vulnerability of MARL MG controllers based on simulations.

A. Simulation Methodology

We consider the microgrid shown in Figure 2, with the parameters given in [20]. We consider that the reference voltage frequency and magnitudes are $\omega_{ref} = 2\pi \times 60$Hz and $v_{ref} = 380$ v, respectively, The utilized base load values were $R_{L1} = 2.5\Omega$, $L_{L1} = 1$mH, $R_{L2} = 3\Omega$, and $L_{L2} = 2$mH. The inverters, network, and load dynamics were simulated according to the differential equations in Section III-A. For secondary control, we considered that the operator uses the MPC controller in (29) [4] for frequency control, and considered two alternatives for secondary control of the voltage magnitude: (i) The MPC controller [4], and (ii) MARL controller trained on the POSG defined in Section III. For the MPC controller, we considered two variants. The first variant is the MPC controller (i.e., (30)) with the the adjacency matrix A of the communication graph and other control parameters as specified in [25]. The second variant is the robust extension proposed by [17], which discards neighbouring DG measurements that are too high or too low. The extension (refered to as MPC-R) requires the communication graph to be "r-robust", and that the number of attacked agents $|\mathcal{G}^a|$ be known *a priori* to the controller. Our considered MG does not satisfy r-robustness, hence MPC-R corresponds to fully distributed MPC (i.e., based on local measurements only). For the MARL controllers, we also considered two variants: the first variant was trained using the proximal policy optimization (PPO) [24] algorithm, the second using the advantage actor-critic (A2C) [26] algorithm. Both algorithms are based on the actor-critic architecture, and represent the state-of-the-art in solving RL problems for environments with continuous action spaces, which is the case in our considered POSG M_c. To implement PPO and A2C, we used the multi-agent versions of the algorithms in the RL-lib Python library [27] using the default algorithm parameters in RL-lib. We experimented with other state-of-the-art RL algorithms, such as soft actor critic (SAC) and deep deterministic policy gradient (DDPG), but those did not yield satisfactory results despite initial attempts for hyper-parameter tuning in RL-lib.

We trained the PPO and A2C controllers for 25,000 episodes each, with an episode length of 1.2 seconds, as this was a sufficient time for stabilizing the voltage in the non-attacked case. During the first 0.2 seconds of an episode, only primary control is active and the secondary control references are computed as $(\omega_{ni} = \omega_{ref}, V_{ni} = v_{ref})$. The secondary controller becomes active after $t = 0.2$ seconds. The sampling time was $T_s = 0.005$ s, thus each episode includes 240 samples. To simulate varying loads in each episode, we considered a mean load resistance and inductance per episode that varies within $\pm 20\%$ of the base values (R_{L1}, L_{L1}, R_{L2}, L_{L2}) mentioned earlier. Within each episode, the instantaneous loads vary uniformly within $\pm 5\%$ of the mean values.

We evaluated the MPC and MARL controllers under three scenarios: First, a system without attacks, i.e., $a^\omega_i = a^v_i = 0, \forall i \in \mathcal{G}$. Second, a random attack where measurement perturbations follow a uniform distribution $a^\omega_i \sim \mathcal{U}(0, a^\omega_{max}), a^v_i \sim \mathcal{U}(0, a^v_{max}), \forall i \in \mathcal{G}^a$. Third, an optimal, RL-based attack, where (a^ω_i, a^v_i) are computed as the actions of the POMDP given by the tuple M_a in Section IV. To train the RL attacks, we utilized the single-agent PPO implementation in the RL-lib Python library, trained for 20,000 episodes. Unless otherwise specified, we considered the following: attacks target all agents (i.e., $\mathcal{G}^a = \mathcal{G}^e = \mathcal{G}$), $a^\omega_{max} = 0.01$ p.u. (i.e., $2\pi \times$ (0.6 Hz)), and $a^v_{max} = 0.1$ p.u. (i.e., 38 Volts).

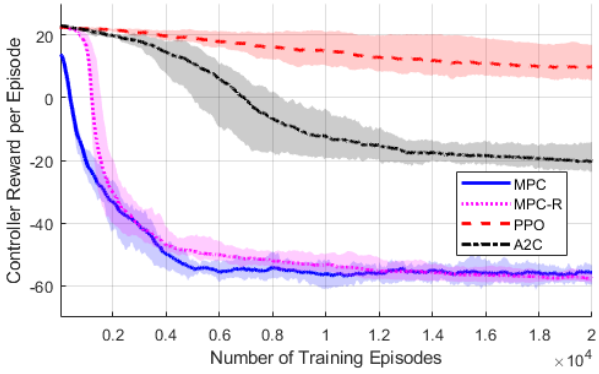

Figure 4: Training curves for the RL attacks against the PPO, A2C and MPC controller.

B. Training Curves

Figure 4 shows the training curves of the RL-based attacks (RLAs) against the three considered MG secondary controllers, obtained as the average of five independently trained agents. The bold lines show the mean reward per episode and the shaded regions show the maximum and minimum rewards among the trained agents. We can observe that the training of all four attacks is very stable, as the variation between the trained agents is small, especially towards the end of the training. The results also show that the MPC and MPC-R controllers are more vulnerable to attacks than the MARL controllers, both in terms of the controller reward and in terms of the number of training episodes required to train a powerful attack. This can be explained by the fact that the output of MARL controllers are more stochastic compared to the MPC controllers, which makes learning an optimal attack against MARL controllers more difficult. Among the two MARL controllers, the PPO controller proves to be more resilient, both in terms of attack impact and in terms of the required attack training duration. This is consistent with results obtained on Atari2600 games in [28], where trust region policy optimization (TRPO), an algorithm from the same family as PPO, was shown to be more resilient to attacks than A3C, which in turn is similar to A2C.

C. Attack Impact on Voltage Stability

Figure 5 shows the maximum voltage deviation Δ^v_{max} achieved under the three attack scenarios, obtained as the average of the maximum voltage deviation across the 4 DGs over 50 episodes. The RLA results were obtained using the best performing agent among the five agents trained (c.f., Figure 4). The results confirm the results in Figure 4 in that the MARL controllers (and especially the PPO controller) are more resilient to FDIAs than the model-based controllers. Interestingly, random attacks against the MPC controller can cause more damage to the MG than the optimal RL attacks against the MARL controllers, highlighting the resilience of MARL-based control. Furthermore, the results show a significant difference between the impact of random attacks and that of RL attacks, which highlights the importance of using an RL attack for vulnerability and impact assessment.

Moreover, Figure 6 shows the the maximum voltage deviation Δ^v_{max} caused by the attacks (in p.u.) as a function of the maximum

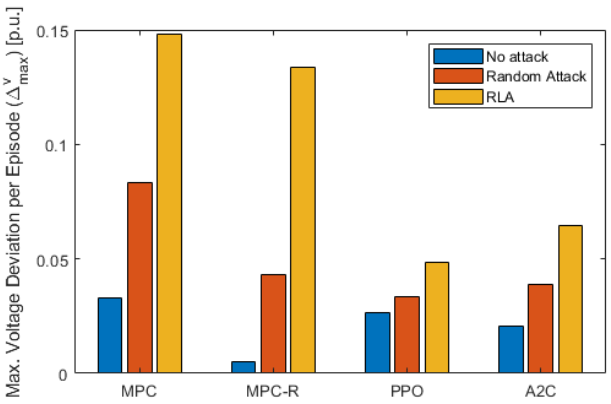

Figure 5: Maximum voltage deviation achieved by the attacks against the MPC, PPO and A2C controllers.

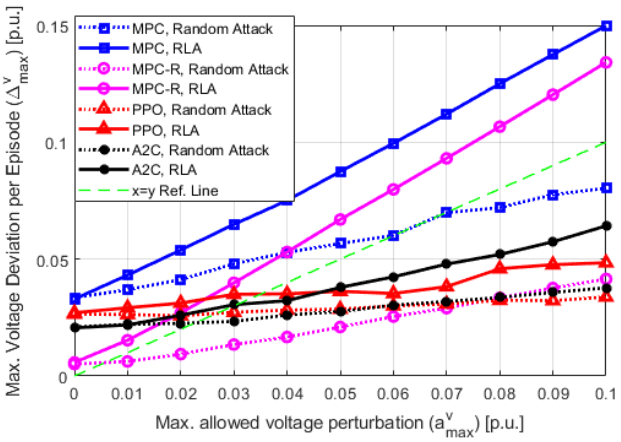

Figure 6: Maximum voltage deviations achieved by the attacks against the considered controllers, as a function of the attack intensity.

attack budget a^v_{max}, for a value of $a^\omega_{max} = 0.01$ p.u. Each curve shows the average of 50 episodes. Somewhat surprisingly, the results show that the attack impact increases approximately linearly with the maximum allowed voltage perturbation under both attack scenarios. Compared to the $x = y$ line, the figure also shows that the marginal gain of increased voltage perturbation of an RL attack against the MPC and MPC-R controllers is always above one, but this is not the case for MARL controllers. Thus, the MARL controllers are overall more robust to voltage perturbation attacks.

D. Impact of Attack Model

In what follows we investigate the impact of reduced attacker capabilities and limited information availability on the vulnerability of the controllers.

1) Attacking a Few Agents Would Suffice: Figure 7 shows the maximum voltage deviation Δ^v_{max} caused by the attacks (in p.u.) as a function of the number of attacked DGs (i.e., $|\mathcal{G}^a|$). Each curve shows the average of 50 episodes. Importantly, the figure shows that the attack impact is concave in the number of attacked DGs, $|\mathcal{G}^a|$. For almost all considered scenarios, an attack affecting only two DGs can achieve almost the same impact as an attack against all DGs. These results further illustrate the threat of FDIAs against MGs, as an attacker with limited access to DGs can inflict significant impact on the MG.

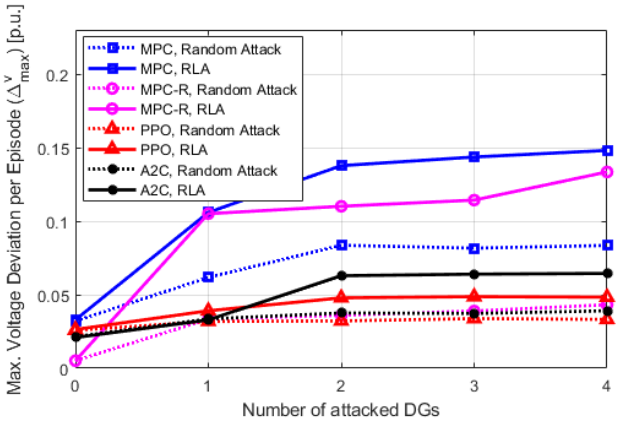

Figure 7: Maximum voltage deviations achieved by the attacks against the considered controllers, as a function of the number of attacked agents (DRs).

		Test Controller ($\mathcal{C}$)			
		MPC	MPC-R	PPO	A2C
Training Controller ($\mathcal{C}'$)	MPC	**0.166**	0.133	0.066	0.089
	MPC-R	0.169	**0.133**	0.064	0.087
	PPO	0.143	0.105	**0.048**	0.064
	A2C	0.144	0.114	0.056	**0.068**

Table I: Voltage deviation Δ^v_{max} caused by an RL attack trained for secondary controller $\mathcal{C}'$ when applied to secondary controller $\mathcal{C}$. Attacker can manipulate the measurements of all DGs.

Furthermore, the figure confirms the clear superiority of the MARL controllers compared to the MPC controller in terms of resiliency, irrespective of the number of attacked DGs.

2) Attacks Trained with Inaccurate Controller Model May Be More Impactful: Finally, we evaluate the transferability of the RL attack across controllers, i.e., how does an RL attack trained for a controller $\mathcal{C}'$ perform when the operator uses controller $\mathcal{C}$. The results in Table I show the attack impact, with the diagonal values (in bold) corresponding to the case ($\mathcal{C}' = \mathcal{C}$). Surprisingly, the attacks with $\mathcal{C}' = \mathcal{C}$ are not always the most impactful. The results clearly show that attacks trained against the MPC and MPC-R controllers are more impactful than other attacks, even when the employed controller is a MARL controller. We attribute this phenomenon to that the response of the MPC and MPC-R controllers is smoother and less stochastic compared to that of the MARL controllers, which facilitates the training of the RL attack.

Overall, these results show that even attacks trained using incomplete information could pose a significant threat to the stability of microgrids. Moreover, the results confirm the superior resilience of the MARL controllers, which is evident when comparing values in the same row of the table. (i.e., the impact achieved by attacks trained on the same $\mathcal{C}'$). We can thus conclude that employing MARL for secondary voltage control in microgrids could be a significant step towards more attack-resilient MGs.

VI. Conclusion

In this paper, we conducted the first study on the vulnerability of MARL-based microgrid secondary control to false data injection attacks. We formulated the problem of MG secondary control as a POSG, and formulated the problem of computing FDIAs against secondary control as a POMDP, and used state-of-the-art RL algorithms to solve the two problems. Our results indicate that traditional model-based secondary controllers are more vulnerable to FDIAs than MARL controllers,

even when the attacker knowledge or capabilities in terms of the number of compromised DGs are limited. Our results show that MARL controllers are significantly more resilient to FDIAs than model-based controllers, and thus could constitute a significant step towards more attack-resilient MGs. Potential future work could further investigate the resilience of MG secondary control by (i) considering different MG topologies, (ii) comparing a broader set of control algorithms, and (iii) evaluating the performance of different FDIA detection schemes.

References

[1] A. Bidram, A. Davoudi, and F. L. Lewis, "A multiobjective distributed control framework for islanded AC microgrids," *IEEE Trans. on Ind. Informatics*, vol. 10, no. 3, pp. 1785–1798, 2014.

[2] J. Lopes, C. Moreira, and A. Madureira, "Defining control strategies for microgrids islanded operation," *IEEE Trans. on Power Systems*, vol. 21, no. 2, pp. 916–924, 2006.

[3] Q. Shafiee, J. M. Guerrero, and J. C. Vasquez, "Distributed secondary control for islanded microgrids—a novel approach," *IEEE Trans. on Power Electronics*, vol. 29, no. 2, pp. 1018–1031, 2014.

[4] G. Lou, W. Gu, Y. Xu, M. Cheng, and W. Liu, "Distributed MPC-based secondary voltage control scheme for autonomous droop-controlled microgrids," *IEEE Trans. on Sustainable Energy*, vol. 8, no. 2, pp. 792–804, 2017.

[5] H. Zhang, W. Meng, J. Qi, X. Wang, and W. X. Zheng, "Distributed load sharing under false data injection attack in an inverter-based microgrid," *IEEE Trans. on Industrial Electronics*, vol. 66, no. 2, pp. 1543–1551, 2019.

[6] A. O. Aluko, R. P. Carpanen, D. G. Dorrell, and E. E. Ojo, "Real-time cyber attack detection scheme for standalone microgrids," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 481–21 492, 2022.

[7] S. Abhinav, H. Modares, F. L. Lewis, F. Ferrese, and A. Davoudi, "Synchrony in networked microgrids under attacks," *IEEE Trans. on Smart Grid*, vol. 9, no. 6, pp. 6731–6741, 2018.

[8] D. Chen, K. Chen, Z. Li, T. Chu, R. Yao, F. Qiu, and K. Lin, "Powernet: Multi-agent deep reinforcement learning for scalable powergrid control," *IEEE Trans. on Power Systems*, vol. 37, no. 2, pp. 1007–1017, 2022.

[9] J. Lin, K. Dzeparoska, S. Q. Zhang, A. Leon-Garcia, and N. Papernot, "On the robustness of cooperative multi-agent reinforcement learning," in *Proc. of IEEE Security and Privacy Workshops (SPW)*, 2020, pp. 62–68.

[10] Y. Li and J. Yan, "Cybersecurity of smart inverters in the smart grid: A survey," *IEEE Trans. on Power Electronics*, vol. 38, no. 2, pp. 2364–2383, 2023.

[11] M. Chlela, D. Mascarella, G. Joós, and M. Kassouf, "Fallback control for isochronous energy storage systems in autonomous microgrids under denial-of-service cyber-attacks," *IEEE Trans. on Smart Grid*, vol. 9, no. 5, pp. 4702–4711, 2018.

[12] S. Liu, X. Wang, and P. X. Liu, "Impact of communication delays on secondary frequency control in an islanded microgrid," *IEEE Trans. on Ind. Electronics*, vol. 62, no. 4, pp. 2021–2031, 2015.

[13] I. Zografopoulos and C. Konstantinou, "Detection of malicious attacks in autonomous cyber-physical inverter-based microgrids," *IEEE Trans. on Industrial Informatics*, vol. 18, no. 9, pp. 5815–5826, 2022.

[14] A. Mustafa, B. Poudel, A. Bidram, and H. Modares, "Detection and mitigation of data manipulation attacks in AC microgrids," *IEEE Trans. on Smart Grid*, vol. 11, no. 3, pp. 2588–2603, 2020.

[15] M. Ma, A. Lahmadi, and I. Chrisment, "Detecting a stealthy attack in distributed control for microgrids using machine learning algorithms," in *Proc. of IEEE Conf. on Industrial Cyberphysical Systems (ICPS)*, vol. 1, 2020, pp. 143–148.

[16] X. Li, C. Wen, C. Chen, and Q. Xu, "Adaptive resilient secondary control for microgrids with communication faults," *IEEE Trans. on Cybernetics*, vol. 52, no. 8, pp. 8493–8503, 2022.

[17] A. Bidram, B. Poudel, L. Damodaran, R. Fierro, and J. M. Guerrero, "Resilient and cybersecure distributed control of inverter-based islanded microgrids," *IEEE Trans. on Industrial Informatics*, vol. 16, no. 6, pp. 3881–3894, 2020.

[18] R. Yan, Y. Wang, Y. Xu, and J. Dai, "A multiagent quantum deep reinforcement learning method for distributed frequency control of islanded microgrids," *IEEE Trans. on Control of Network Systems*, vol. 9, no. 4, pp. 1622–1632, 2022.

[19] N. Pogaku, M. Prodanovic, and T. C. Green, "Modeling, analysis and testing of autonomous operation of an inverter-based microgrid," *IEEE Trans. on Power Electronics*, vol. 22, no. 2, pp. 613–625, 2007.

[20] A. Bidram, A. Davoudi, F. L. Lewis, and J. M. Guerrero, "Distributed cooperative secondary control of microgrids using feedback linearization," *IEEE Trans. on Power Systems*, vol. 28, no. 3, pp. 3462–3470, 2013.

[21] M. H. Khooban and M. Gheisarnejad, "A deep reinforcement learning controller based type-II fuzzy system: Frequency regulation in microgrids," *IEEE Trans. on Emerging Topics in Computational Intelligence*, vol. 5, no. 4, pp. 689–699, 2021.

[22] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 ukraine blackout: Implications for false data injection attacks," *IEEE Trans. on Power Systems*, vol. 32, no. 4, pp. 3317–3318, 2017.

[23] B. Li, R. Lu, G. Xiao, H. Bao, and A. Ghorbani, "Towards insider threats detection in smart grid communication systems," *IET Communications*, vol. 13, 07 2019.

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

[25] D. Chen, "Large-scale islanded microgrids based on pid control methods," https://github.com/DongChen06/Microgrid, 2020.

[26] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Harley, T. P. Lillicrap, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. of Intl. Conf. on Machine Learning*, 2016, p. 1928–1937.

[27] E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, and I. Stoica, "RLlib: Abstractions for distributed reinforcement learning," in *Proc. of Intl. Conf. on Machine Learning*, 2018, pp. 3053–3062.

[28] S. Huang, N. Papernot, I. Goodfellow, Y. Duan, and P. Abbeel, "Adversarial attacks on neural network policies," *arXiv preprint arXiv:1702.02284*, 2017.