Securing Cyber-Physical LFC Systems: A Resilient MADDPG Approach Under FDI Attacks

Anoosh Dini
Electrical Engineering
Polytechnique Montreal
Montreal, Canada
anoosh.dini@polymtl.ca

Keyhan Sheshyekani
Electrical Engineering
Polytechnique Montreal
Montreal, Canada
keyhan.sheshyekani@polymtl.ca

Hanane Dagdougui
Math. and Industrial Engineering
Polytechnique Montreal
Montreal, Canada
hanane.dagdougui@polymtl.ca

***Abstract*—This paper investigates the resilience of multi-agent deep deterministic policy gradient (MADDPG)-based automatic generation control (AGC) systems against false data injection (FDI) attacks. The vulnerability of this control method to small perturbations in agents' observations caused by FDI attacks is first analyzed. Then, we propose a retraining phase for the standard MADDPG in which gradient-based perturbations are injected into each agent's observations, making the model resilient against FDI attacks. The perturbations are bounded by a tunable radius to maintain nominal performance and do not alter the decentralized execution architecture. Simulation results on the IEEE 39-bus system indicate that the proposed method effectively limits frequency deviations under both attack and no-attack conditions.**

***Index Terms*—Cyber-security, deep learning, load-frequency control, power system dynamics, reinforcement learning.**

I. Introduction

Frequency is a critical factor in the stability of power grids. Whenever it deviates from its normal value, controllers must act quickly to restore the balance. Load-frequency control (LFC) accomplishes this task with the help of automatic generation control (AGC). The AGC measures and calculates the area control error (ACE) and adjusts the setpoints of generation units to restore frequency to its nominal value [1]. The quality of the communication between the sensors, the controller, and the generators plays a direct role in the effectiveness of this process [2].

In recent years, extensive research has been conducted on the design of AGC controllers. These studies can be classified into two major groups: classical control-based methods and artificial intelligence (AI)-based approaches [3]. The former is either based on metaheuristic algorithms to tune the coefficients of the proportional-integral-derivative (PID) controllers or relies on other classical control methods, such as fuzzy control or robust control, to fulfill specific goals. Such goals may include maintaining frequency under high penetration of renewable energy resources (RESs) or under uncertainties in grid parameters. However, metaheuristic algorithms are not suitable for AGC applications, which require real-time gain tuning [3]. Furthermore, approaches such as fuzzy controllers only perform well under the scenario conditions for which they are tuned, so they may otherwise have oscillatory responses [4]. These challenges have led to a shift towards AI-based AGC techniques, which provide enhanced performance in large-scale interconnected systems with RES integration alongside real-time decision-making. Although these schemes can be trained using historical data or through interaction with simulation models, they are fundamentally based on deep neural networks (DNNs).

For example, authors in [5] propose a deep reinforcement learning (DRL) method with a discrete action space. However, this method does not perform well due to the limitations caused by the discreteness of the controller's action space. Moreover, adding more neural networks to the DRL architecture to make the model more complex, as is done in [4], does not necessarily improve the performance of the controller. Instead, it makes training more difficult and adds to the number of hyperparameters that require fine-tuning. Compared with previous studies, [6] adopts the deep deterministic policy gradient (DDPG) method, whose action space is continuous. However, because the training and execution of controllers are decentralized, each controller operates independently. Therefore, the proposed approach cannot effectively handle tie-line power exchanges between neighboring areas. To resolve this issue, [7] develops a multi-agent DDPG (MADDPG) approach, where inter-area interactions are modeled via centralized training and decentralized execution. The effectiveness of MADDPG for AGC applications has also been confirmed in [8], [9].

As discussed, the LFC system is a combination of physical and cyber aspects. However, what studies such as [7], [8], [9] have in common is that they focus only on controlling the physical layer while neglecting the cyber layer and its associated challenges. A significant challenge at this layer is false data injection (FDI) attacks, in which attackers attempt to disrupt system stability by manipulating data. In LFC systems, data is transmitted through the communication infrastructure, which connects measurement sensors, the control center (i.e., the controller), and actuators. As communication between the sensors and the control center is less likely to be encrypted, this link becomes particularly vulnerable to FDI attacks. Countermeasures against such attacks are typically implemented at the control center using state estimation and bad data detection (BDD) algorithms [2]. However, these methods cannot detect all attacks with certainty [10]. As a result, the controller may

979-8-3315-8138-1

make control actions based on manipulated data. Since the performance of the MADDPG method and, more broadly, DRL algorithms depends on the accuracy of the agents' observations (i.e., measurement data), it is essential to make these algorithms resilient to such attacks.

To highlight the research gap, the cyber layer is generally neglected in the implementation of the MADDPG-based AGC. As previously discussed, FDI attacks directly affect the measurement data based upon which the DDPG agent determines setpoints for generation units. Since the agent only acts on what it observes and does not have access to the real state of the grid, even small changes can provide a false image of the system conditions and may lead to the system instability [11]. In particular, FDI attacks on frequency or tie-lines power measurements may alter the setpoints provided by AGC while evading detection by conventional BDD methods. Based on the identified research gap, this study aims to investigate the impact of FDI attacks on MADDPG performance. Later, we develop a resilient version of the algorithm to ensure stability against FDI attacks.

II. System Model and Performance

A. LFC Model

The block diagram of the LFC model for a control area in a multi-area power system is shown in Fig. 1. The physical dynamics match the standard LFC structure. To model the physical system, we adopt the standard small-signal LFC model around an operating point. For area $i \in \{1, \dots, N\}$,

$$M_i \frac{d\Delta f_i}{dt} = \Delta P_{m,i} - \Delta P_{L,i} - \Delta P_{tie,i} - D_i \Delta f_i \tag{1}$$

$$T_{gi} \frac{d\Delta P_{g,i}}{dt} + \Delta P_{g,i} = \Delta P_{c,i} - \frac{1}{R_i} \Delta f_i \tag{2}$$

$$T_{ti} \frac{d\Delta P_{m,i}}{dt} + \Delta P_{m,i} = \Delta P_{g,i} \tag{3}$$

$$\frac{d\Delta P_{tie,i}}{dt} = 2\pi \sum_{j=1, j\neq i}^{N} T_{ij} (\Delta f_i - \Delta f_j). \tag{4}$$

Here, Δf_i denotes the frequency deviation of area i, $\Delta P_{m,i}$ and $\Delta P_{g,i}$ represent the mechanical and governor power deviations, respectively, and $\Delta P_{tie,i}$ is the deviation in tie-line power flow between area i and its neighboring areas. The parameters M_i, D_i, T_{gi}, and T_{ti} correspond to the inertia constant, damping coefficient, governor time constant, and turbine time constant, respectively. Furthermore, $\Delta P_{L,i}$ represents the change in load demand. Constants R_i and β_i are the droop characteristic of the governor and the frequency bias factor, respectively.

In the cyber layer, which is responsible for AGC, the ACE signal is calculated by

$$\text{ACE}_i = \Delta P_{tie,i} + \beta_i \Delta f_i. \tag{5}$$

This signal is then transmitted to the controller through the communication network. Following this step, the controller generates $\Delta P_{c,i}$ and sends it to the governor in the physical layer. Since the data transmission path involves components such as sensors and communication links, it is susceptible to FDI attacks. Such attacks can introduce errors or deviations in ACE. Although the path between the control center and the generation units is typically encrypted, the measurement path remains particularly vulnerable to cyber threats.

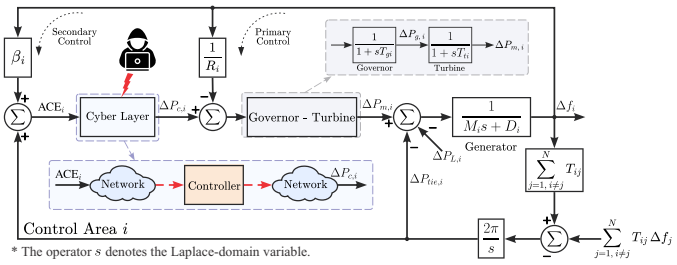

Fig. 1. Block diagram of the cyber-physical LFC model of a control area in an interconnected power system.

B. MADDPG Scheme

This subsection presents the MADDPG-based AGC framework, adapted from [7]. As shown in Fig. 2, the AGC problem is formulated as a cooperative Markov game under the centralized training and decentralized execution framework. In this regard, each controller is represented by an independent agent that is trained by a centralized critic. Each agent includes an actor and a critic, both of which have online and target neural networks to ensure stable convergence. Online actor network and online critic network are denoted by $\mu_i(o_i; \theta_i^\mu)$ and $Q_i(\boldsymbol{s}, \boldsymbol{a}; \phi_i^Q)$, respectively, and their corresponding target networks are expressed by μ_i' and Q_i'. Here, i-th agent's observation is represented by o_i, and the global state vector $\boldsymbol{s}$ includes all agents' observations. Similarly, the joint action vector can be expressed by $\boldsymbol{a} = (a_1, \dots, a_N)$. As shown in Fig. 2, the critic has access to $\boldsymbol{a}$ and $\boldsymbol{s}$, allowing it to properly estimate the value function by considering the interactions among agents. In addition, each agent learns to improve its action based solely on its own observation. The online critic network is trained to minimize the loss function defined by

$$\mathcal{L}_i(\phi_i^Q) = \mathbb{E}\Big[\big(y_i - Q_i(\boldsymbol{s}_t, \boldsymbol{a}_t; \phi_i^Q)\big)^2\Big] \tag{6}$$

where the target value y_i is computed using the target networks

$$y_i = r_{i,t} + \gamma\, Q_i'\big(\boldsymbol{s}_{t+1}, \boldsymbol{a}'_{t+1}; \phi_i^{Q'}\big). \tag{7}$$

Here, $r_{i,t}$ is the reward for agent i at time t, and γ is the discount factor. The target action $\boldsymbol{a}'_{t+1}$ is obtained by passing the next observations through the target actor networks of all agents, as given by

$$\boldsymbol{a}'_{t+1} = \big(\mu_1'(o_{1,t+1}), \dots, \mu_N'(o_{N,t+1})\big). \tag{8}$$

The actor for each agent is trained by applying the policy gradient to maximize the expected return from its own perspective, as expressed by

$$\nabla_{\theta_i^\mu} \mathcal{J}_i = \mathbb{E}\Big[\nabla_{a_i} Q_i\big(\boldsymbol{s}_t, \boldsymbol{\mu}(\boldsymbol{o}_t); \phi_i^Q\big)\ \nabla_{\theta_i^\mu} \mu_i(o_{i,t}; \theta_i^\mu)\Big]. \tag{9}$$

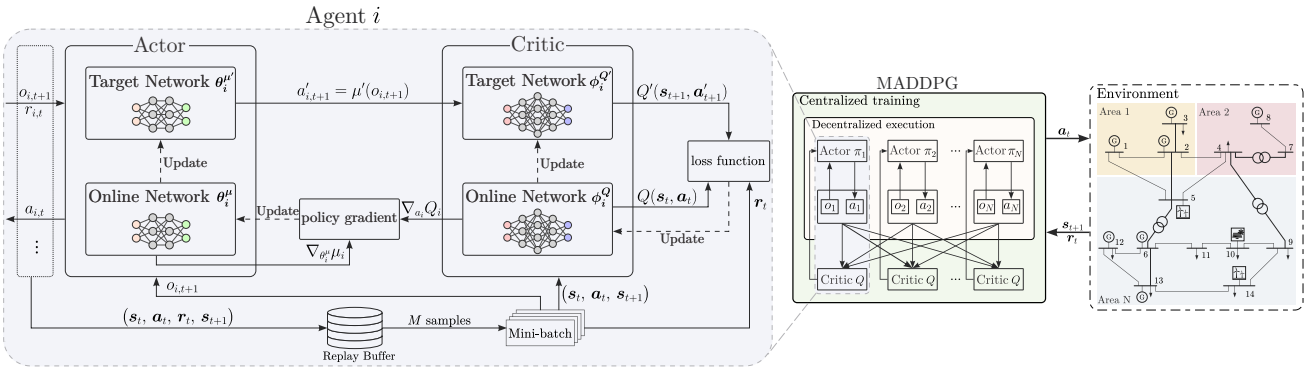

Fig. 2. Overview of the MADDPG scheme including centralized training and decentralized execution for AGC of a multi-area power system.

Equation (9) can be interpreted as the actor receiving feedback from the critic, indicating how to adjust its output action so that the expected value estimated by the critic increases. During training, the online actor and online critic networks are updated sequentially (the critic first, then the actor). In contrast, updating the parameters of the target networks follows the rule $\phi_i^{Q'} \leftarrow \tau\phi_i^{Q} + (1-\tau)\phi_i^{Q'}$ and $\theta_i^{\mu'} \leftarrow \tau\theta_i^{\mu} + (1-\tau)\theta_i^{\mu'}$, where $\tau \ll 1$ is the soft-update coefficient.

C. Vulnerability Analysis

The execution phase of the MADDPG framework is decentralized; otherwise, each agent would require access to the global system state, without which it could not determine the control action. Moreover, during decentralized execution, each agent is independently responsible for maintaining the frequency stability of its own area, similar to the operation of conventional LFC controllers. Therefore, during execution, all networks are discarded except the online actor network, which determines the control action based on the agent's local observations. An attacker can induce an undesirable control action by manipulating measurements that form part of the agent's local observations, as shown in Fig. 3. As a result, the frequency stability of the system might be disrupted.

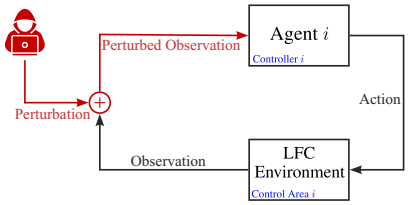

Fig. 3. Attacker targets the decentralized execution of agent i.

Any perturbation vector, δ_i, can disrupt the agent's understanding of the true environment conditions. In this case, the agent does not accurately perceive the true state o_i. Instead, it observes a perturbed state $\hat{o}_i$, given by $\hat{o}_i = o_i + \delta_i$. Since the agent is not aware of this perturbation, it determines new setpoints for generation units based on the manipulated data, which results in a suboptimal action denoted by $\hat{a}_i = \mu_i(\hat{o}_i; \theta_i^{\mu})$. The difference between the desired action and the performed action is defined by $\Delta a_i = \hat{a}_i - a_i$. By taking the first-order Taylor expansion of the actor network around the true observation o_i, it can be shown that for small perturbations δ_i, the change in the control action is linearly related to the size of the perturbation

$$\Delta a_i \approx J_\mu(o_i)\, \delta_i \tag{10}$$

where $J_\mu(o_i) = \nabla_{o_i} \mu_i(o_i; \theta_i^{\mu})$ is the Jacobian matrix of the actor network's output with respect to its input. The vulnerability of the model to observation deviation from its true value is highlighted in (10). Since the norm of this Jacobian can be very large for a DNN [4], the effect of δ_i can be even amplified by this value.

III. Resilient Model for MADDPG

In this section, a training mechanism is proposed that enhances the resilience of the MADDPG framework against FDI attacks through a gradient-based perturbation procedure. In the proposed method, the standard MADDPG model is first trained on attack-free observations to ensure the model's baseline control performance. Then, a retraining phase is performed for all agents using perturbed observations that simulate malicious manipulation of the measurements.

In standard MADDPG, the critic network updates the actor network in a way that the actor selects actions with the highest expected return. This is achieved by updating the actor in the direction of the gradient $\nabla_{a_i} Q_i$. In contrast, here we seek the change in the agent's observation that achieves the highest expected return, given by the gradient $\nabla_{o_i} Q_i$. Therefore, the worst-case disturbance for MADDPG is the perturbation to the observation that most decreases the critic's value at the current state (i.e., it moves along the negative of this gradient, $-\nabla_{o_i} Q_i$) [11]. Finding the worst-case perturbation directions and retraining the model based on them is inspired by principles of robust control. As a result of this retraining phase, actor and critic networks gradually learn to take actions that remain effective even when the measurements are compromised. The overall strategy can be viewed as a min-max optimization problem in which each agent seeks an optimal action that maximizes its expected return under the worst allowable perturbation in its local observations.

At each iteration of the retraining phase, the true observation o_i of agent i is first perturbed by the gradient of the critic's value function with respect to the agent's observation. This perturbation is generated as $\delta_i = -\epsilon \frac{\nabla_{o_i} Q_i(\boldsymbol{s},\boldsymbol{a})}{\|\nabla_{o_i} Q_i(\boldsymbol{s},\boldsymbol{a})\|}$, where ϵ is the perturbation radius. As discussed, the gradient points in the direction in which a small change in the observation significantly reduces the estimated Q-value.

Algorithm 1 Proposed resilient MADDPG

Require: N agents; online actor μ_i and target actor μ_i'; online critic Q_i and target critic Q_i'; replay buffer $\mathcal{D}$; mini-batch $\mathcal{M}$; discount factor γ; soft update τ; perturbation radius ϵ; retraining episodes K, length T.

Phase I: Standard MADDPG training

Train standard MADDPG on clean observations

Phase II: Resilient MADDPG retraining

for $k = 1, \ldots, K$ **do**
  Reset environment; observe initial global state $\boldsymbol{s}$
  for $t = 1, \ldots, T$ **do**
    for $i = 1, \ldots, N$ **do**
      $a_i \leftarrow \mu_i(o_i)$
      $\delta_i \leftarrow -\epsilon \frac{\nabla_{o_i} Q_i(\boldsymbol{s},\boldsymbol{a})}{\|\nabla_{o_i} Q_i(\boldsymbol{s},\boldsymbol{a})\|}$
      $\hat{o}_i = o_i + \delta_i$
      $\hat{a}_i \leftarrow \mu_i(\hat{o}_i)$
    end for
    Apply joint action $\hat{\boldsymbol{a}}_t = (\hat{a}_{1,t}, \ldots, \hat{a}_{N,t})$; receive rewards $\boldsymbol{r}_t$ and next state $\boldsymbol{s}_{t+1}$.
    Store tuple $(\boldsymbol{s}_t, \hat{\boldsymbol{a}}_t, \boldsymbol{r}_t, \boldsymbol{s}_{t+1})$ in $\mathcal{D}$
    Sample mini-batch $\mathcal{M} \subset \mathcal{D}$
    for all $(\boldsymbol{s}_t, \hat{\boldsymbol{a}}_t, \boldsymbol{r}_t, \boldsymbol{s}_{t+1}) \in \mathcal{M}$ **do**
      for $i = 1, \ldots, N$ **do**
        $a'_{i,t+1} \leftarrow \mu'_i(o_{i,t+1})$
      end for
      Form $\boldsymbol{a}'_{t+1} = (a'_{1,t+1}, \ldots, a'_{N,t+1})$
      for $i = 1, \ldots, N$ **do**
        Set $y_i \leftarrow \boldsymbol{r}_t + \gamma\, Q'_i(\boldsymbol{s}_{t+1}, \boldsymbol{a}'_{t+1})$
      end for
    end for
    for $i = 1, \ldots, N$ **do**
      Update online critic by replacing $\boldsymbol{a}_t$ with $\hat{\boldsymbol{a}}_t$ in (6).
      Update online actor using (9)
      Soft-update target networks
    end for
  end for
end for

Execution: Each agent i applies $a_i = \mu_i(o_i)$ in decentralized operation.

IV. Simulation Studies

To investigate the impact of FDI attacks on the standard MADDPG and to evaluate the performance of the proposed resilient model under such attacks, the IEEE 39-bus system is used as the test case. The system consists of nine generators and three control areas. It is a common benchmark for LFC studies, and its parameters are provided in [1]. According to the proposed method, each area's controller is considered as an agent, and all agents are trained centrally based on the standard MADDPG model. In this case, $o_i = [\Delta f_i, \Delta P_{tie,i}, \Delta P_{m,i}, \int \text{ACE}_i, \frac{d}{dt}\text{ACE}_i]$ defines the observation of agent i, based on which the actor network of each agent determines the control action during the decentralized execution phase. The centralized training process is carried out over 1000 episodes with a simulation time step of 0.1 s. During each episode, intentional load disturbances are applied in all control areas to ensure that the agents learn their optimal control actions. Each agent's actor network maps the observation state o_i to a control action u_i. The centralized critic network takes all the observations (o_1, o_2, o_3) and actions (u_1, u_2, u_3) from all three agents as its input to estimate the Q value. The model uses a discount factor (γ) of 0.995 and a soft target update rate (τ) of 0.003. The actor network learning rate is set to 0.0001, while the critic network learning rate is 0.001. Experiences are stored in a replay buffer with a capacity of 10^6 transitions, and network updates are performed using a minibatch size of 256. To evaluate the standard MADDPG in the presence of FDI attacks, we assume that the first area is subject to random perturbations $\delta_1 \sim \mathcal{N}(0, 0.02^2)$, which only affect the first two elements of o_1. According to [10], larger perturbations are more likely to be detected by BDD.

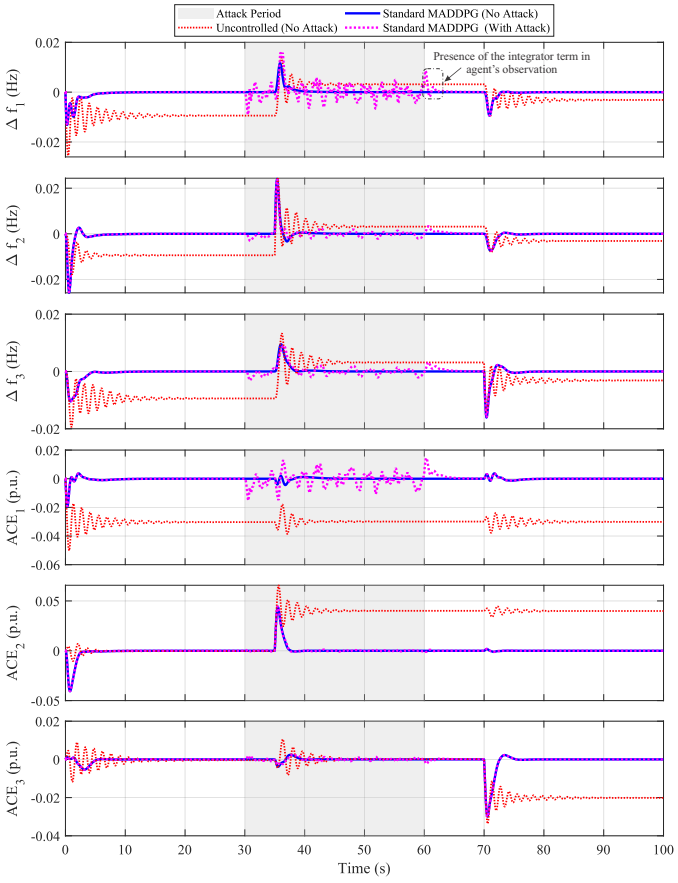

Fig. 4. Random perturbations affecting Δf_1 and $\Delta P_{tie,1}$.

Moreover, the first two elements of o_1 are targeted because they are associated with the sensor measurements.

As shown in Fig. 4, the controller performance in the first area is severely affected; consequently, the other two areas are also impacted. Although the attack ends at 60 seconds, the integrator term (i.e., $\int \text{ACE}_i$) in the agent observations causes the attack's effect to remain for several seconds.

To enhance the model's resilience against attacks, the MADDPG is retrained for 100 episodes according to the second phase of Algorithm 1. In this case, all controllers

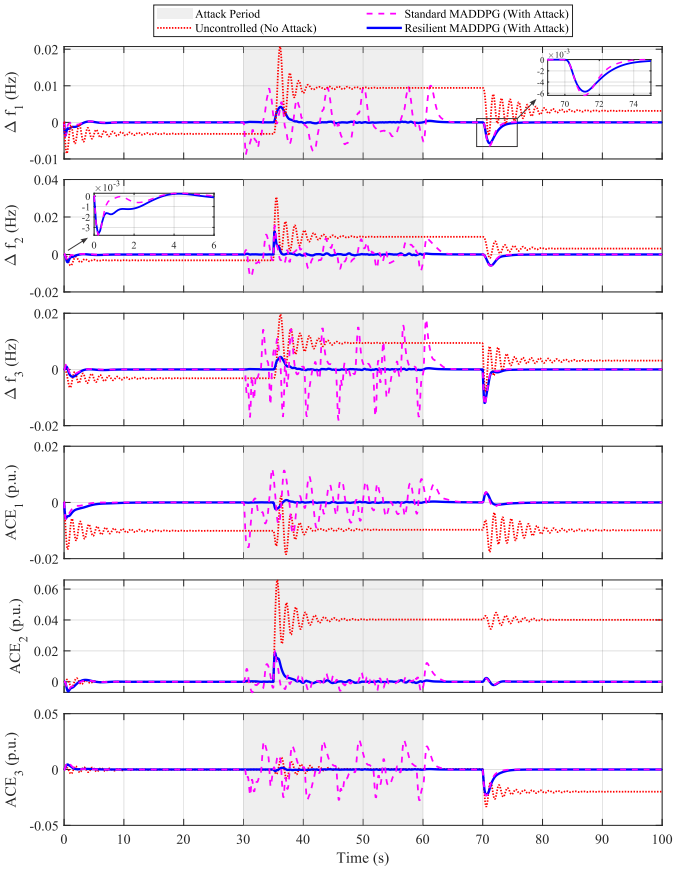

Fig. 5. Evaluation of the proposed resilient MADDPG under simultaneous attacks perturbing the observations of agents.

in the three areas are simultaneously subjected to random perturbations following the same normal distribution as before. As shown in Fig. 5, the proposed model maintains acceptable resilience against attacks on agent observations. In addition, it is important to note that the number of retraining episodes and the perturbation radius have significant impacts on the model's performance under no-attack conditions. Therefore, these parameters should be selected such that the control actions under no-attack conditions remain close to those of the standard model, as shown in Fig. 5.

V. Conclusion

In this work, we study the resilience of MADDPG-based AGC under FDI attacks and demonstrate that even small perturbations in agent observations can lead to harmful control actions. To address this vulnerability, a resilient retraining mechanism was introduced for the standard MADDPG framework, in which gradient-based perturbations are applied during learning. Simulation on the IEEE 39-bus system showed that the proposed approach enhances frequency recovery and minimizes ACE under FDI attacks. Moreover, the model maintains near-baseline control performance under normal operating conditions when the perturbation radius is properly tuned.

References

[1] H. Bevrani, *Robust power system frequency control*. Springer, 2009.

[2] A. M. Mohan, N. Meskin, and H. Mehrjerdi, "A comprehensive review of the cyber-attacks and cyber-security on load frequency control of power systems," *Energies*, vol. 13, no. 15, p. 3860, 2020.

[3] Ibraheem, P. Kumar, and D. Kothari, "Recent philosophies of automatic generation control strategies in power systems," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 346–357, 2005.

[4] 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, 2020.

[5] D. Zhang et al., "Research on AGC performance during wind power ramping based on deep reinforcement learning," *IEEE Access*, vol. 8, pp. 107 409–107 418, 2020.

[6] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1653–1656, 2019.

[7] Z. Yan and Y. Xu, "A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4599–4608, 2020.

[8] R. Muduli, D. Jena, and T. Moger, "Application of reinforcement learning-based adaptive PID controller for AGC of multi-area power system," *IEEE Trans. Automa. Sci. Eng.*, vol. 22, pp. 1057–1068, 2025.

[9] J. Li, T. Yu, and X. Zhang, "Coordinated load frequency control of multi-area integrated energy system using multi-agent deep reinforcement learning," *Applied energy*, vol. 306, p. 117 900, 2022.

[10] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[11] A. Pattanaik, Z. Tang, S. Liu, G. Bommannan, and G. Chowdhary, "Robust deep reinforcement learning with adversarial attacks," *arXiv preprint arXiv:1712.03632*, 2017.