Towards Learning-Enhanced Distributed Resilient and Safe Control of DC Microgrids Under False Data Injection Attacks

Mohamadamin Rajabinezhad, Alaa Selim, Suining He, Junbo Zhao, and Shan Zuo

***Abstract*— This paper proposes a fully–distributed Attack-Resilient and Safety-Aware (DARSA) control framework for DC microgrids subject to diverse false data injection (FDI) attacks on control input channels. Unlike existing approaches that treat transient-stat safety and steady-state resilience as separate objectives, DARSA ensures both simultaneously by maintaining bus voltages within prescribed safety bounds even at attack onset. This capability is crucial for low-inertia converter-based microgrids, which are highly vulnerable to severe overshoots and fluctuations during transients triggered by disturbances or cyber intrusions. The framework adopts a consensus-based secondary control strategy with adaptive coupling gains to mitigate FDI impacts. Safety and convergence are certified through CBF-based constraints and Lyapunov stability concepts, ensuring voltage regulation and uniformly ultimately bounded (UUB) performance. An ensemble of machine-learning (ML) surrogate models is further used to efficiently tune controller parameters under a limited simulation budget. Simulation studies and hardware-in-the-loop validation using Typhoon HIL demonstrate improved reliability, safety, and resilience under a broad class of FDI attacks.**

I. INTRODUCTION

Distributed control of DC microgrids has attracted significant attention due to its scalability, reliability, and resilience to single points of failure compared with centralized schemes. Key objectives include maintaining the average bus voltage and achieving proportional load sharing among converters [1]. These goals are commonly addressed through a hierarchical framework: primary droop control provides fast regulation, while distributed secondary control restores voltage deviations with low communication overhead [1], [2]. However, the tight coupling of secondary control with communication networks exposes microgrids to cyber threats such as false-data injection (FDI), denial-of-service (DoS), and replay attacks, which can disrupt voltage regulation and load sharing [3]. Ensuring both transient stability and transient safety is therefore critical, yet maintaining states within safe bounds during dynamics remains less explored. Existing work primarily mitigates cyber–physical attacks through detection or resilient control. Detection approaches often require restrictive assumptions on the number of compromised nodes or network connectivity [4], whereas resilient control compensates attacks in real time without explicit identification. Safety is especially important in low-inertia microgrids, where cyber-induced transients can induce harmful voltage excursions [5]. Although Model predictive control (MPC) can enforce safety constraints, its computational burden limits real-time deployment. Control barrier functions (CBFs), widely used for safety-critical control [6], offer an efficient alternative.

This paper introduces a novel fully distributed Attack-Resilient and Safety-Aware (DARSA) secondary control framework for DC microgrids, designed to withstand diverse range of FDI attacks. The key contributions of this work are:

M. Rajabinezhad, S. He and S. Zuo are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA (email: mohamadamin.rajabinezhad@uconn.edu, suining.he@uconn.edu, shan.zuo@uconn.edu). A. Selim and J. Zhao are with Thayer School of Engineering, Dartmouth College, Hanover, NH, 0377, USA. (email: alaa.e.selim@dartmouth.edu, junbo.zhao@dartmouth.edu).

1) We propose a fully distributed DARSA framework that ensures voltage safety and resilient operation of DC microgrids under diverse FDI attacks. Unlike prior methods limited to steady-state performance [3], [7], DARSA combines adaptive time-varying coupling gains with CBF-based safety constraints and Lyapunov analysis to guarantee UUB voltage regulation and stable load sharing, mitigating transient overshoots caused by low inertia and cyber disturbances.
2) We provide a Lyapunov-based analysis establishing uniformly ultimately bounded (UUB) resilience and integrate an ensemble machine-learning surrogate for sample-efficient controller tuning under constrained simulation resources, with simulation and HIL validation confirming improved stability, safety, and resilience under unbounded FDI attacks.

Notations and Graph Theory: $\mathbf{1}_N$ denotes the all-ones vector in $\mathbb{R}^N$, $|\cdot|$ the absolute value, and $\mathrm{diag}\{\cdot\}$ the diagonal map. The DC microgrid communication graph is $\mathscr{G} = (\mathscr{W}, \mathscr{E}, \mathcal{A})$ with nodes $\mathscr{W} = \{0, 1, \ldots, N\}$ (leader 0). Its adjacency matrix $\mathcal{A} = [a_{ij}]$ satisfies $a_{ij} > 0$ iff $(w_j, w_i) \in \mathscr{E}$. The neighbor set is $\mathcal{N}_i = \{j : a_{ij} > 0\}$, the in-degree matrix $\mathcal{D} = \mathrm{diag}(d_i)$ with $d_i = \sum_{j\in\mathcal{N}_i} a_{ij}$, the Laplacian $\mathcal{L} = \mathcal{D} - \mathcal{A}$, and the pinning matrix $\mathcal{G} = \mathrm{diag}(g_i)$ with $g_i > 0$ for pinned nodes.

Preliminaries: CBF and ISSf–CBF. Consider the affine system

$$\dot{x} = f(x) + g(x)u, \tag{1}$$

with locally Lipschitz f, g, state $x \in \mathbb{R}^n$, and input $u \in \mathbb{R}^m$.

Definition 1 ([6]). *A set Ω is forward invariant for* (1) *if $x(t_0) \in \Omega$ implies $x(t) \in \Omega$ for all $t \geq t_0$.*

Definition 2 (CBF [6]). *Let $\Omega = \{x \mid h(x) \geq 0\}$ for $h \in C^1$. Then h is a control barrier function if there exists a class-$\mathcal{K}$ function α such that* $\sup_{u\in\mathbb{R}^m}\{L_f h(x) + L_g h(x)u + \alpha(h(x))\} \geq 0$.

Lemma 1. *[[6]] Let h be a CBF for system* (1) *and $\Omega = \{x \in \mathbb{R}^n : h(x) \geq 0\}$. Any locally Lipschitz feedback $u(x)$ satisfying $L_f h(x) + L_g h(x)u(x) + \alpha(h(x)) \geq 0$ for all x renders Ω forward invariant.*

Assume the applied input is affected by a bounded signal $\mu(t)$, so $u \mapsto u + \mu(t)$ with $\|\mu\|_\infty \leq \bar{\mu}$. Following the Input-to-state safety (ISSf) [8], formulation for sets, we seek forward invariance of the *inflated* set

$$C_d(\bar{\mu}) \;:=\; \big\{\, x : \; h(x) + \gamma(\bar{\mu}) \;\geq 0 \,\big\}, \tag{2}$$

for some class-$\mathcal{K}$ function γ. Since $\gamma(\bar{\mu}) \geq 0$, we have $C \subseteq C_d(\bar{\mu})$; moreover, if $\bar{\mu}_1 \leq \bar{\mu}_2$ then $C_d(\bar{\mu}_1) \subseteq C_d(\bar{\mu}_2)$, and $C_d(\bar{\mu}) \to C$ as $\bar{\mu} \to 0$, i.e., C_d expands with disturbance size and collapses to C as the disturbance vanishes [8].

Definition 3 (ISSf–BF [8]). *h is an ISSf–BF on $\mathcal{D} \supset C$ if there exist class-$\mathcal{K}$ functions α, ι such that $L_f h(x) + \alpha(h(x)) \;\geq\; -\iota(|\mu|)$, for all $x \in \mathcal{D}$ and bounded μ.*

Definition 4 (ISSf–CBF [8]). *h is an ISSf–CBF if there exist class-$\mathcal{K}$ functions α, ι such that*

$$\sup_{u\in U} \big[L_f h(x) + L_g h(x)(u + \mu)\big] \geq -\alpha(h(x)) - \iota(|\mu|). \tag{3}$$

979-8-3315-8138-1

Lemma 2. *[8]. For a continuously differentiable function h defined in* (3)*, any locally Lipschitz selection $u(x) \in U$ that satisfies the inequality renders $C_d(\|\mu\|_\infty)$ in* (2) *forward invariant.*

II. Cyber-Physical DC Microgrids

The autonomous DC microgrid (Fig. 2) uses four voltage-controlled buck converters. Secondary cooperative control regulates bus voltages and enforces proportional sharing via exchange of $\mathrm{X}_i = [\bar{V}_i,\ R_i^{\mathrm{vir}} I_i]$, where $\bar{V}_i$ is the average-voltage estimate, I_i the output current, and $R_i^{\mathrm{vir}} = k/I_i^{\mathrm{rated}}$. Consensus on $R_i^{\mathrm{vir}} I_i$ provides proportional load sharing [2], [7]. Primary droop shapes V_i^*, and the secondary layer updates V_{n_i}. The local voltage setpoint satisfies

$$V_i^* = V_{n_i} + V_{\mathrm{ref}} - R_i^{\mathrm{vir}} I_i, \tag{4}$$

and differentiating (4) yields the linearized secondary input

$$\dot{V}_{n_i} = \dot{V}_i^* + R_i^{\mathrm{vir}} \dot{i}_i = u_i. \tag{5}$$

The cooperative secondary controller is

$$u_i^c = c_i \left(g_i (V_{\mathrm{ref}} - \bar{V}_i) + \sum_{j \in \mathcal{N}_i} a_{ij} (R_j^{\mathrm{vir}} I_j - R_i^{\mathrm{vir}} I_i) \right), \tag{6}$$

with coupling gain $c_i > 0$. The global average voltage estimate evolves as

$$\bar{V}_i(t) = V_i(t) + \int \sum_{j \in \mathcal{N}_i} a_{ij} \big(\bar{V}_j(t) - \bar{V}_i(t) \big)\, dt. \tag{7}$$

Assuming $V_i^* = V_i$, combining (5)–(7) yields

$$\begin{aligned}\dot{\bar{V}}_i + R_i^{\mathrm{vir}} \dot{i}_i = c_i \Big(&\sum_{j \in \mathcal{N}_i} a_{ij} \big[(\bar{V}_j + R_j^{\mathrm{vir}} I_j) - (\bar{V}_i + R_i^{\mathrm{vir}} I_i) \big] \\ &+ g_i \big[(V_{\mathrm{ref}} + R_i^{\mathrm{vir}} I_i) - (\bar{V}_i + R_i^{\mathrm{vir}} I_i) \big] \Big).\end{aligned} \tag{8}$$

Following [1], define $\Theta_i = \bar{V}_i + R_i^{\mathrm{vir}} I_i, \Theta_{\mathrm{ref}} = V_{\mathrm{ref}} + k I_{ss}^{\mathrm{pu}}$, leading to

$$\dot{\Theta}_i = c_i \left(\sum_{j \in \mathcal{N}_i} a_{ij} (\Theta_j - \Theta_i) + g_i (\Theta_{\mathrm{ref}} - \Theta_i) \right). \tag{9}$$

Its compact global form is

$$\dot{\Theta} = -\mathrm{diag}(c_i)(\mathcal{L} + \mathcal{G})(\Theta - \mathbf{1}_N \Theta_{\mathrm{ref}}). \tag{10}$$

where $\Theta = \left[\Theta_1^{\mathrm{T}}, \ldots, \Theta_N^{\mathrm{T}} \right]^{\mathrm{T}}$. Define the cooperative regulation error

$$\varepsilon = \Theta - \mathbf{1}_N \Theta_{\mathrm{ref}}, \tag{11}$$

where $\varepsilon = \left[\varepsilon_1^{\mathrm{T}}, \ldots, \varepsilon_N^{\mathrm{T}} \right]^{\mathrm{T}}$, which satisfies

$$\dot{\varepsilon} = -\mathrm{diag}(c_i)(\mathcal{L} + \mathcal{G})\varepsilon. \tag{12}$$

Equations (10)–(12) show that consensus-based secondary control achieves global voltage regulation and proportional load sharing over a connected communication graph.

Assumption 1. *The communication network $\mathscr{G}$ is an undirected and connected graph.*

Lemma 3 ([3]). *Given Assumption 1, $(\mathcal{L}+\mathcal{G})$ is nonsingular and positive-definite.*

Lemma 4 ([2]). *Under Assumption 1, designing the auxiliary control input as in* (6) *and* (7) *ensures that both global voltage regulation and proportional load sharing are achieved.*

Malicious actors may inject false data into the local control input of each converter. Thus, instead of (5), the compromised secondary input becomes

$$\dot{V}_{n_i} = \dot{V}_i^* + R_i^{\mathrm{vir}} \dot{I}_i = \bar{u}_i = u_i + \delta_i(t), \tag{13}$$

where $\bar{u}_i(t)$ is the corrupted input and $\delta_i(t)$ denotes the unknown injection, which may be constant, time-varying, bounded, or unbounded.

Assumption 2. *The p-th order derivative ($p > 0$) of the locally injected false data is bounded, i.e., there exists $\bar{\kappa} > 0$ such that $\|\delta_i^{(p)}\| \leq \bar{\kappa}$.*

Remark 1. *Most secondary-level works consider only bounded FDI signals or bounded derivatives [3], [7]. Real attacks may grow or follow nonlinear patterns (e.g., $t \sin t$, polynomials), potentially driving $\dot{V}_{n_i}$ unbounded. Assumption 2 accommodates such cases while keeping the analysis manageable; the parameter p trades off attack coverage and complexity.*

Consider the attack injections in (13). Under the conventional secondary control (6)–(7), substituting into (12) yields the perturbed error dynamics

$$\dot{\varepsilon} = -\,\mathrm{diag}(c_i)(\mathcal{L} + \mathcal{G})\varepsilon + \delta, \tag{14}$$

where $\delta = [\delta_1^{\mathrm{T}}, \ldots, \delta_N^{\mathrm{T}}]^{\mathrm{T}}$ is the attack vector. If δ is unbounded, then ε becomes unbounded as well, indicating that the standard secondary controller cannot ensure stability under general FDI disturbances. This motivates the need for an attack-resilient design. For convergence analysis, we adopt the following definition.

Definition 5 ([2]). *$x(t) \in \mathbb{R}$ is UUB with the ultimate bound b, if there exist constants $b, c > 0$, independent of $t_0 \geq 0$, and for every $a \in (0, c)$, there exists $t_1 = t_1(a, b) \geq 0$, independent of t_0, such that $|x(t_0)| \leq a \Rightarrow |x(t)| \leq b, \quad \forall t \geq t_0 + t_1$.*

A safe voltage region can be prescribed around the nominal operating point, and the converter voltages must remain within this region throughout operation, including under FDI attacks. Each DG is required to satisfy the voltage bounds $[V_l, V_h]$, determined by hardware limits and transient safety considerations. The constraint can be written as

$$-V_l \leq \tilde{V}_i \leq V_h, \tag{15}$$

where $\tilde{V}_i = V_i - V_{\mathrm{nom}}$. In the presence of FDI disturbances, the goal is to design u_i such that all voltages remain within the safety set and converge to a neighborhood of the leader-issued reference. This leads to the following attack-resilient, safety-aware control problem for DC microgrids.

Problem (Distributed Attack-Resilient and Safety-Aware (DARSA) Control Problem)**.** *Under the unknown diverse FDI attacks on local control input channels, design local control protocols u_i in* (13) *for each converter, to ensure that (i) the safety constraints specified in* (15) *are guaranteed, (ii) ε in* (12) *is UUB.* That is, bounded global voltage regulation and proportional load sharing are both achieved.

III. Fully–Distributed DARSA Controller Design

A. Safety Guarantees Using CBFs

To enforce the voltage bounds in (15), define the CBFs as $h_{V_{1,i}} = \tilde{V}_i + V_l, h_{V_{2,i}} = V_h - \tilde{V}_i$, so safety is equivalent to $h_{V_{1,i}} \geq 0$ and $h_{V_{2,i}} \geq 0$, yielding the set $\Omega_{V_i} = \{\tilde{V}_i \mid h_{V_{1,i}} \geq 0,\ h_{V_{2,i}} \geq 0\}$. Since $\dot{h}_{V_{1,i}} = \dot{V}_i$ and $\dot{h}_{V_{2,i}} = -\dot{V}_i$, we employ linear class-$\mathcal{K}$ functions $\alpha_1 = \eta_1 h_{V_{1,i}}, \alpha_2 = \eta_2 h_{V_{2,i}}$, which give the zeroing-CBF conditions

$$\dot{h}_{V_{1,i}} + \eta_1 h_{V_{1,i}} \geq 0, \qquad \dot{h}_{V_{2,i}} + \eta_2 h_{V_{2,i}} \geq 0. \tag{16}$$

Substituting $\dot{V}_i$ from (5) introduces the unknown term $R_i^{\mathrm{vir}} \dot{I}_i$. As load variations are finite, we assume $|\dot{I}_i| \leq d_I$ [9].

ISS-tightened CBF constraints. Let the control channel contain a bounded perturbation $\mu_i(t)$ with $\|\mu_i\|_\infty \leq \bar{\mu}_i$. Since $L_g h_{V_{1,i}} = 1$ and $L_g h_{V_{2,i}} = -1$, the global slope bound is $\vartheta_{V,i} = 1$, and the ISS tightening yields

$$L_f h_{V,i} + L_g h_{V,i}\, u_i \geq -\alpha(h_{V,i}) + \bar{\mu}_i.$$

979-8-3315-8138-1

Combining the load-rate and input perturbation bounds gives the robust CBF inequalities

$$u_i - R_i^{\mathrm{vir}} d_I - \bar{\rho}_i + \eta_1 (V_i + V_l - V_{\mathrm{nom}}) \geq 0, \quad (17)$$

$$-u_i + R_i^{\mathrm{vir}} d_I - \bar{\rho}_i + \eta_2 (V_h - V_i + V_{\mathrm{nom}}) \geq 0. \quad (18)$$

when $\bar{\rho}_i = 0$ these reduce to the standard robust CBF rows; for $\bar{\rho}_i > 0$ they certify invariance of the inflated set $C_d(\bar{\rho}_i)$ from the preliminaries. To impose safety with minimal deviation from the nominal controller u_i^c, we solve the QP

$$\begin{cases} \arg \min\limits_{u_i^{safe}} \| u_i^{safe} - u_i^c \|_2 \\ \text{s.t.} \quad A_{cbf} u_i^{safe}(t) \leq b_{cbf} \end{cases} \quad (19)$$

where $A_{cbf} = [-1; 1]$ and $b_{cbf} = [\eta_1 (V_i + V_l - V_{nom}) - R_i^{\mathrm{vir}} d_I - \bar{\rho}_i; \eta_2 (V_h - V_i + V_{nom}) + R_i^{\mathrm{vir}} d_I - \bar{\rho}_i]$. Denote $\Delta u_i \equiv u_i^{safe} - u_i^c$. In fact, this quadratic programming (QP) problem can be explicitly solved [9].

B. Fully–distributed exponentially attack-resilient Controller

The formulation and components of the proposed distributed exponentially attack-resilient controller are now presented. Denote $\zeta_i = \sum\limits_{j \in \mathcal{N}_i} a_{ij} (\bar{V}_j - \bar{V}_i) + g_i (V_{\mathrm{ref}} - \bar{V}_i) + \sum\limits_{j \in \mathcal{N}_i} a_{ij} (R_j^{\mathrm{vir}} I_j - R_i^{\mathrm{vir}} I_i)$. To achieve bounded global voltage regulation and proportional load sharing in the presence of FDI attacks, a fully distributed attack-resilient control strategy is introduced, aiming to enhance DC microgrid robustness:

$$u_i = \mathrm{H}(\xi_i) \, u_i^{safe}, \quad (20)$$

$$\dot{\xi}_i = \alpha_i (|\zeta_i| - \beta_i \left(\xi_i - \hat{\xi}_i \right)), \quad (21)$$

$$\dot{\hat{\xi}}_i = \rho_i \left(\xi_i - \hat{\xi}_i \right), \quad (22)$$

where $\mathrm{H}(\xi_i) = (1 + \xi_i)^p$ and $p \in N$, $p \geq 1$, ξ_i is adaptively tuned by (21). α_i, β_i, and ρ_i are positive constants. Next, we give the main result of solving the attack-resilient secondary control problems for DC microgrid. Since FDI attacks and resilient compensation may create residual disturbances in the barrier dynamics, an ISS–CBF is adopted to maintain forward invariance.

Theorem 1. *Given Assumptions 1 and 2, for DC microgrid under the diverse FDI attacks, by applying DARSA control protocols consist of* (7)*,* (20)*,* (21) *and* (22)*, the safety constraints specified in* (15) *are almost guaranteed all the time, and the cooperative regulation error* ε *in* (11) *is UUB. That is, the DARSA control problem is solved.*

Proof: Please see Appendix. ■

IV. Ensemble Machine-Learning-Based Tuning of Controller Parameters

This section presents a data-driven tuning framework for controller parameters based on an ensemble of machine-learning (ML) surrogate models. The approach treats the closed-loop system and its time-domain simulation as a black-box oracle and constructs a regression ensemble that approximates the mapping between controller gains and a scalar performance index. Let $\boldsymbol{\theta} \in \mathbb{R}^d$ denote the vector of controller parameters to be tuned. We assume that a nominal parameter vector $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ is available from prior design or heuristic tuning. To reduce the search space and preserve relative magnitudes, we introduce a dimensionless scaling vector

$$\mathbf{s} = [s_1, \ldots, s_d]^\top \in [s_{\min}, s_{\max}]^d, \quad (23)$$

and define the actual controller parameters as

$$\boldsymbol{\theta}(\mathbf{s}) = \mathrm{diag}(\boldsymbol{\theta}_0) \, \mathbf{s}. \quad (24)$$

For a given parameter vector $\boldsymbol{\theta}$, the time-domain simulation of the closed-loop system produces trajectories $\mathbf{y}(t; \boldsymbol{\theta})$ over a time horizon $[0, T]$. A scalar performance index $J(\boldsymbol{\theta})$ is then computed from these trajectories. In tracking problems, a common choice is a windowed mean-squared error (MSE) with respect to a reference signal $\mathbf{r}(t)$:

$$J(\boldsymbol{\theta}) = \frac{1}{T_{\mathrm{w}}} \int_{t_1}^{t_1 + T_{\mathrm{w}}} \left\| \mathbf{y}(t; \boldsymbol{\theta}) - \mathbf{r}(t) \right\|_2^2 \mathrm{d}t, \quad (25)$$

where t_1 and T_{w} denote the start time and duration of the evaluation window, respectively. More general performance metrics (e.g., weighted sums of MSE, overshoot) can be incorporated into (25) without modifying the proposed tuning pipeline. The tuning objective is to identify a scaling vector $\mathbf{s}^\star$ that minimizes the cost in (25) within a prescribed simulation budget $N_{\max}$:

$$\mathbf{s}^\star = \arg \min_{\mathbf{s} \in [s_{\min}, s_{\max}]^d} J\big(\boldsymbol{\theta}(\mathbf{s})\big). \quad (26)$$

Due to the implicit dependence of J on $\mathbf{s}$ through a nonlinear simulation model, gradients are not readily available, and classical gradient-based tuning is not directly applicable. Given a maximum number of allowable simulations $N_{\max}$, we first allocate $N_{\mathrm{train}} < N_{\max}$ evaluations to construct a training dataset for the surrogate models. The first evaluation is performed at the nominal scaling as $\mathbf{s}^{(1)} = \mathbf{1}_d$, which corresponds to $\boldsymbol{\theta}(\mathbf{s}^{(1)}) = \boldsymbol{\theta}_0$. The remaining $N_{\mathrm{train}} - 1$ points are drawn from a logarithmically uniform design over the hyper-rectangle $[s_{\min}, s_{\max}]^d$, i.e.,

$$\log s_i^{(k)} \sim \mathcal{U}\big(\log s_{\min}, \log s_{\max}\big), i = 1, \ldots, d, \ k = 2, \ldots, N_{\mathrm{train}}. \quad (27)$$

For each scaling vector $\mathbf{s}^{(k)}$, the corresponding controller parameters $\boldsymbol{\theta}(\mathbf{s}^{(k)})$ are injected into the simulation model, the system response $\mathbf{y}(t; \boldsymbol{\theta}(\mathbf{s}^{(k)}))$ is computed, and the cost $J^{(k)} = J(\boldsymbol{\theta}(\mathbf{s}^{(k)}))$ is evaluated using (25). This yields the training dataset

$$\mathcal{D} = \big\{ \big(\mathbf{s}^{(k)}, \, J^{(k)}\big) \, \big| \, k = 1, \ldots, N_{\mathrm{train}} \big\}. \quad (28)$$

On the dataset $\mathcal{D}$, we train a collection of M regression models

$$f_m : [s_{\min}, s_{\max}]^d \to \mathbb{R}, \qquad m = 1, \ldots, M, \quad (29)$$

each approximating the unknown mapping from scaling factors to cost,

$$f_m(\mathbf{s}) \approx J\big(\boldsymbol{\theta}(\mathbf{s})\big). \quad (30)$$

In the proposed framework, f_m may include heterogeneous ML regressors such as: ❶ linear or generalized linear models, ❷ Gaussian Procress Regression (GPR), ❸ support vector regression, ❹ bagged decision trees, ❺ boosted decision trees. To assess the predictive quality of each surrogate, we perform K-fold cross-validation on $\mathcal{D}$ and compute the root-mean-square error (RMSE):

$$\mathrm{RMSE}_m = \sqrt{ \frac{1}{N_{\mathrm{train}}} \sum_{k=1}^{N_{\mathrm{train}}} \big(J^{(k)} - f_m(\mathbf{s}^{(k)}) \big)^2 }. \quad (31)$$

These RMSE values are reported for each model, providing a transparent comparison of their accuracy. We form an ensemble surrogate by aggregating the individual predictors. A simple yet effective choice is the unweighted mean:

$$\widehat{J}_{\mathrm{ens}}(\mathbf{s}) = \frac{1}{M} \sum_{m=1}^{M} f_m(\mathbf{s}), \quad (32)$$

although weighted combinations (e.g., with weights inversely proportional to RMSE_m) can also be employed. After training the ensemble surrogate (32), the remaining simulation budget $N_{\max} - N_{\mathrm{train}}$ is used to refine the tuning. We generate a set of candidate scaling vectors $\big\{ \mathbf{s}_j^{\mathrm{cand}} \big\}_{j=1}^{N_{\mathrm{cand}}} \subset [s_{\min}, s_{\max}]^d$. for example by sampling additional points logarithmically uniformly within the search box or by focusing around the best-performing training samples. For each candidate, the ensemble surrogate predicts the cost:

$$\widehat{J}_{\mathrm{ens}}\big(\mathbf{s}_j^{\mathrm{cand}}\big), \quad j = 1, \ldots, N_{\mathrm{cand}}. \quad (33)$$

979-8-3315-8138-1

We then select the most promising candidate according to the surrogate,

$$j^{\star} = \arg \min_{1 \le j \le N_{\mathrm{cand}}} \widehat{J}_{\mathrm{ens}}\big(\mathbf{s}_j^{\mathrm{cand}}\big), \tag{34}$$

and evaluate its true cost by running the simulation and computing $J(\boldsymbol{\theta}(\mathbf{s}_{j^\star}^{\mathrm{cand}}))$. This evaluation may be appended to the dataset $\mathcal{D}$ and used to re-train the surrogates if additional refinement iterations are desired. Finally, the tuned controller parameters are chosen as $\boldsymbol{\theta}^{\star} = \boldsymbol{\theta}\big(\mathbf{s}^{\star}\big)$, where $\mathbf{s}^{\star}$ corresponds to the scaling vector that achieved the lowest observed cost across all evaluated points (nominal, exploratory samples, and surrogate-guided candidates). The ensemble-based tuner was executed with a budget of

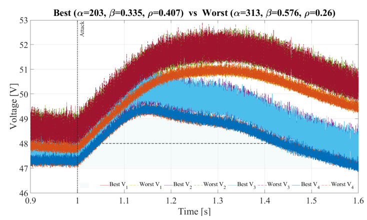

Fig. 1: Voltage signal responses tuned based on Ensemble-ML model.

$N_{\max} = 50$ simulations, where $N_{\mathrm{train}} = 49$ evaluations were used to construct the surrogate and the final simulation was reserved for an ensemble-guided candidate. The design variables are the dimensionless scaling factors $(s_\alpha, s_\beta, s_\rho)$ applied to the nominal gains $(\alpha_0, \beta_0, \rho_0)$, and are constrained to $[s_{\min}, s_{\max}] = [0.8, 1.5]$. During exploration, one of the last purely exploratory samples (evaluation 49) corresponds to $(s_\alpha, s_\beta, s_\rho) = (1.183, 0.9975, 1.105)$ and yields a relatively high cost $J = 3.947$. After training the five surrogate models on the 49 collected samples, the linear model was discarded due to a configuration error, while the remaining four models achieved cross-validated RMSEs of 0.1756 (GPR), 0.7111 (SVM), 0.909 (bagged trees), and 0.4445 (boosted trees). Their averaged prediction was then used to screen 2,000 candidate scaling vectors, and the best ensemble prediction led to the final evaluation $(s_\alpha, s_\beta, s_\rho) = (0.8479, 0.8387, 1.357)$ with $J = 1.594$. Over all 50 simulations, the best observed parameters are $\alpha \approx 203.5$, $\beta \approx 0.335$, and $\rho \approx 0.407$, which produce a mean windowed MSE of 1.59 across the four converters (with individual MSEs of 0.66, 2.47, 2.55, and 0.69 and terminal voltages V_{end} between 47.5–48.5 V). Fig. 1 compares the voltage trajectories under this best setting with those obtained from the worst-performing setting in the search (with $\alpha \approx 313$, $\beta \approx 0.576$, $\rho \approx 0.26$).

V. Validation And Results

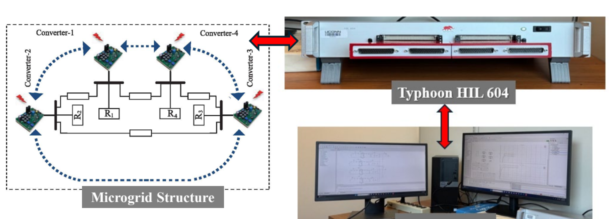

Fig. 2: The tested DC MG physical structure built using Typhoon HIL devices

A low-voltage DC microgrid (Fig. 2) is used for numerical and real-time validation. The system consists of four buck converter–interfaced sources with rated currents $(6, 3, 3, 6)$ A, virtual impedances $(2, 4, 4, 2)\,\Omega$, and parameters $C = 2.2$ mF, $L = 2.64$ mH, $f_s = 60$ kHz, $R_{\mathrm{line}} = 0.1\ \Omega$, $R_L = 10\ \Omega$, $V_{\mathrm{ref}} = 48$ V, and $V_{\mathrm{in}} = 80$ V. Simulations are performed in MATLAB/Simulink with voltage safety bounds $V_l = V_h = 1.5$ V, while real-time validation is conducted on a Typhoon HIL 604 platform using tighter bounds $V_l = V_h = 0.5$ V and a fixed step size of 0.5 μs. The bidirectional communication network yields an undirected Laplacian structure; in the HIL setup, inter-agent communication uses Typhoon's internal synchronous data exchange, yielding delay-free interactions without external protocols. Multiple test cases assess safety and resilience under unbounded FDI attacks.

A. Simulation-Based Evaluation Under FDI Attacks

This case study evaluates the proposed distributed resilient controller under diverse FDI attacks over a 0–10 s interval, with attacks injected at $t = 1$ s as $\delta_i = [0.4t^2 + 5,\ \ 5t\sin t + 7,\ \ 2t + 3,\ \ 0.3t^2 + 12]^\top,\ i = 1, \ldots, 4$. The proposed controller is first evaluated without safety constraints and compared with a conventional distributed secondary controller. As shown in Fig. 3, the conventional controller becomes unstable after attack onset, while the proposed resilient controller maintains bounded voltage regulation and proportional current sharing. However, without explicit safety enforcement, transient voltage excursions violate the prescribed safety limits, indicating that resilience alone is insufficient to guarantee safe operation. Next, the full DARSA controller integrates CBF-based safety constraints into the resilient framework. As shown in Fig. 4, all converter voltages remain within the predefined safety region while achieving UUB convergence in voltage regulation and current sharing, including during attack onset and recovery. This comparison confirms the effectiveness of the unified safety–resilience design for reliable operation under FDI attacks.

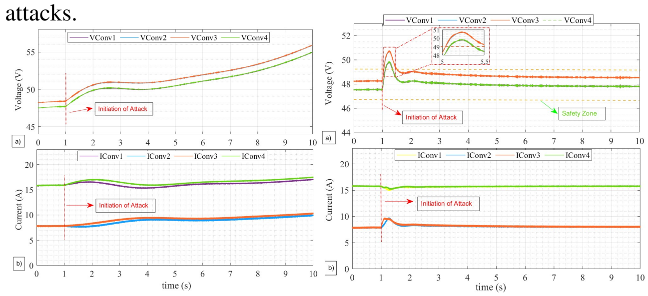

Fig. 3: Comparative performance of the conventional (left) and proposed resilient (right) controllers without safety constraints under unbounded attacks: (a) terminal voltages; (b) supplied currents

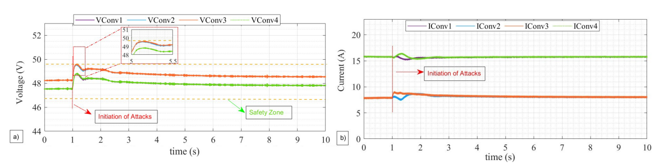

Fig. 4: Simulation performance of the Proposed DARSA controller, in the case of unbounded attack: a) Terminal voltages of Converters , b) Supplied currents.

B. Hardware-in-the-Loop Validation Using Typhoon HIL-604

To further validate the practical feasibility of the proposed framework under realistic real-time constraints, hardware-in-the-loop (HIL) experiments are conducted using the Typhoon HIL-604+ platform. Unbounded FDI attacks are injected into the control input channels at $t = 5$ s using linearly increasing signals, $\delta_i = [0.4t + 130,\ \ 0.3t + 90,\ \ 0.3t + 80,\ \ 0.35t + 100]^\top,\ i = 1, \ldots, 4$. Fig. 5 compares HIL performance of the proposed resilient controller without safety constraints (left) and the full DARSA controller with CBF-based safety enforcement (right) under unbounded FDI attacks. Without safety constraints, voltage regulation and current

979-8-3315-8138-1

sharing remain bounded, but transient voltage excursions violate the admissible safety region following attack onset. With the full DARSA controller, all voltages are maintained within the prescribed safety bounds while preserving regulation and proportional load sharing throughout the experiment. These HIL results are consistent with the simulation studies and confirm the necessity of the unified safety–resilience design for real-time operation under unbounded FDI attacks.

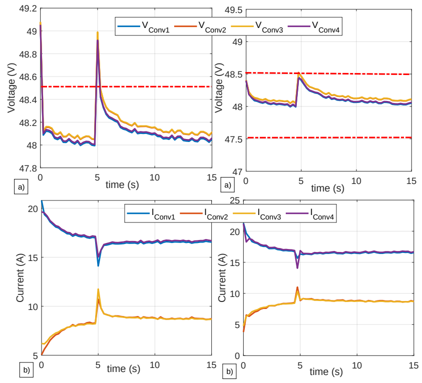

Fig. 5: Comparative HIL performance of the proposed resilient controller without (left) and with (right) safety constraints under unbounded attacks: (a) terminal voltages; (b) supplied currents.

VI. Conclusion

This paper proposes a fully distributed DARSA control framework for DC microgrids under general FDI attacks on control inputs. DARSA ensures transient safety and steady-state resilience by enforcing voltage bounds and guaranteeing UUB convergence in regulation and load sharing. The framework integrates consensus-based secondary control with smart adaptive coupling gains, CBF-based safety enforcement, Lyapunov analysis, and an ensemble machine-learning surrogate for efficient parameter tuning. Both simulation studies and hardware-in-the-loop validation using the Typhoon HIL platform demonstrate improved reliability, safety, and resilience under a broad class of FDI attacks—capabilities not achieved in prior studies.

References

[1] V. Nasirian, S. Moayedi, A. Davoudi, and F. L. Lewis, "Distributed cooperative control of dc microgrids," *IEEE Trans. Power Electronics*, vol. 30, no. 4, pp. 2288–2303, 2014.

[2] Y. Zhang, M. Rajabinezhad, Y. Wang, J. Zhao, and S. Zuo, "Privacy-enhanced defense control framework for dc microgrid against exponentially unbounded attacks," *IEEE Access*, 2025.

[3] S. Zuo, O. A. Beg, F. L. Lewis, and A. Davoudi, "Resilient networked ac microgrids under unbounded cyber attacks," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 3785–3794, 2020.

[4] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2016.

[5] M. Rajabinezhad, N. Shams, Y. Wang, and S. Zuo, "Privacy-preserving, safety-aware, and attack-resilient distributed cooperative control in ac microgrids against exponentially unbounded fdi attacks," *IEEE Trans. Industry Applications*, 2025.

[6] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Trans. Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2016.

[7] P. S. Tadepalli, D. Pullaguram, and M. Alam, "Resilient dynamic average secondary control for dc microgrids against fdi attacks," *IEEE Trans. Industry Applications*, 2025.

[8] S. Kolathaya and A. D. Ames, "Input-to-state safety with control barrier functions," *IEEE control systems letters*, vol. 3, no. 1, pp. 108–113, 2018.

[9] J. Zhang, L. Ding, X. Lu, and W. Tang, "A novel real-time control approach for sparse and safe frequency regulation in inverter intensive microgrids," *IEEE Trans. Industry Applications*, 2023.

[10] J. LaSalle, "Some extensions of liapunov's second method," *IRE Trans. circuit theory*, vol. 7, no. 4, pp. 520–527, 1960.

Proof: We leverage the CBF framework from Theorem 2 in [8] for safety analysis. Defining the safety set Ω_{Vi} with barrier functions $h_{V_{1,i}}$ and $h_{V_{2,i}}$, and applying Lemma 1 and 2, we ensure forward invariance by satisfying (16). This forms the safety constraints in (19), guaranteeing that the QP solution u_i^{safe} maintains non-negative barrier functions, ensuring $\tilde{V}_i \in \Omega_{Vi}$. Using (5), (7) and (13) the global form of time derivative of ζ_i is $\dot{\zeta} = -\left(\mathcal{L}+\mathcal{G}\right)\left(\operatorname{diag}(1+\xi_i)^p\right)\left(\zeta+\Delta u_f\right)+\delta)$, where $\Delta u_f = \left[\Delta u_{f_1}, \cdots, \Delta u_{f_N}\right]^{\mathrm{T}}$. Consider the following Lyapunov function candidate $E\left(t\right) = \frac{1}{2}\sum_{i=1}^{N}\zeta_i^2(1+\xi_i)^p$. Its time derivative along the system trajectory of ζ is given by

$$
\begin{aligned}
\dot{E} =& \sum_{i=1}^{N}\left(\zeta_i\dot{\zeta}_i(1+\xi_i)^p + \frac{p}{2}\zeta_i^2(1+\xi_i)^{p-1}\dot{\xi}_i\right)\\
=& -\zeta^{\mathrm{T}}\operatorname{diag}(1+\xi_i)^p(\mathcal{L}+\mathcal{G})\left(\operatorname{diag}(1+\xi_i)^p\zeta+\delta\right)\\
&-\zeta^{\mathrm{T}}\operatorname{diag}(1+\xi_i)^p(\mathcal{L}+\mathcal{G})\left(\operatorname{diag}(1+\xi_i)^p\Delta u_f\right)\\
&+\frac{p}{2}\zeta^{\mathrm{T}}\operatorname{diag}((1+\xi_i)^{p-1})\operatorname{diag}\left(\dot{\xi}_i\right)\zeta
\end{aligned}
\tag{35}
$$

Based on the norm bound property, for (35), by using Young's inequality $2ab \leq qa^2 + \frac{1}{q}b^2$ where q is a positive number, and Sylvester's inequality $0 \leq \sigma_{\min}(P)\left\|x\right\|^2 \leq x^{\mathrm{T}}Px \leq \sigma_{\max}(P)\left\|x\right\|^2$, where x is a non-zero vector, and P is a positive-definite matrix. Let $W \triangleq \operatorname{diag}\left((1+\xi_i)^p\right)$ and $W_{p-1} \triangleq \operatorname{diag}\left((1+\xi_i)^{p-1}\right)$. one has

$$
\begin{aligned}
\dot{E} \leq& -\sigma_{\min}(\mathcal{L}+\mathcal{G})\left\|W\zeta\right\|^2 + \sigma_{\max}(\mathcal{L}+\mathcal{G})\left\|W\zeta\right\|\left\|\delta\right\|\\
&+\sigma_{\max}(\mathcal{L}+\mathcal{G})\left\|W\zeta\right\|\left\|W\Delta u_f\right\| + \frac{p}{2}\left\|W_{p-1}\zeta\right\|\left\|\operatorname{diag}(\dot{\xi}_i)\zeta\right\|,
\end{aligned}
$$

Assume there exist constants $0 < m_1 \leq M_1 < \infty$ such that, for all time, $m_1 \leq 1+\xi_i \leq M_1$. Since $1+\xi_i > 0$, the matrix $W_1 \triangleq \operatorname{diag}(1+\xi_i)$ is invertible. Then $\|W_{p-1}\zeta\| \leq \|W_1^{-1}\|\,\|W\zeta\| = \frac{1}{m_1}\|W\zeta\|$.

$$
\begin{aligned}
\dot{E} \leq& -\sigma_{\min}(\mathcal{L}+\mathcal{G})\|W\zeta\|^2 + \sigma_{\max}(\mathcal{L}+\mathcal{G})\|W\zeta\|\,\|\delta\|\\
&+\sigma_{\max}(\mathcal{L}+\mathcal{G})\|W\zeta\|\,\|W\Delta u_f\| + \frac{p}{2m_1}\|W\zeta\|\,\|\operatorname{diag}(\dot{\xi}_i)\zeta\|\\
\leq& -\sigma_{\min}(\mathcal{L}+\mathcal{G})\|W\zeta\| \times \Bigg(\|W\zeta\| - \frac{\sigma_{\max}(\mathcal{L}+\mathcal{G})}{\sigma_{\min}(\mathcal{L}+\mathcal{G})}\|\delta\|\\
&-\frac{\sigma_{\max}(\mathcal{L}+\mathcal{G})}{\sigma_{\min}(\mathcal{L}+\mathcal{G})}\|W\Delta u_f\| - \frac{p}{2m_1\sigma_{\min}(\mathcal{L}+\mathcal{G})}\|\operatorname{diag}(\dot{\xi}_i)\|\,\|\zeta\|\Bigg)
\end{aligned}
\tag{36}
$$

Given Assumption 1, $\sigma_{\min}\left(\mathcal{L}+\mathcal{G}\right) > 0$. Then, a sufficient condition for $\dot{E}(t) \leq 0$ is

$$
\begin{aligned}
&(1+\xi_i)^p\left|\zeta_i(t)\right| - \frac{\sigma_{\max}\left(\mathcal{L}+\mathcal{G}\right)}{\sigma_{\min}\left(\mathcal{L}+\mathcal{G}\right)}(1+\xi_i)^p\left|\Delta u_{f_i}\right|\\
&-\frac{\sigma_{\max}\left(\mathcal{L}+\mathcal{G}\right)}{\sigma_{\min}\left(\mathcal{L}+\mathcal{G}\right)}\left|\delta_i(t)\right| - \frac{p\left|\dot{\xi}_i(t)\right|\left|\zeta_i(t)\right|}{2m_1\sigma_{\min}\left(\mathcal{L}+\mathcal{G}\right)} \geq 0.
\end{aligned}
\tag{37}
$$

Define $\tilde{\xi}_i(t) = \xi_i(t) - \hat{\xi}_i(t)$, based on (21) and (22), its time derivative is

$$
\begin{aligned}
\dot{\tilde{\xi}}_i(t) =& \alpha_i\left|\zeta_i(t)\right| - (\alpha_i\beta_i+\rho_i)\left(\xi_i(t) - \hat{\xi}_i(t)\right)\\
=& \alpha_i\left|\zeta_i(t)\right| - (\alpha_i\beta_i+\rho_i)\tilde{\xi}_i(t).
\end{aligned}
\tag{38}
$$

For (38), based on the integral solution for differential equations, one has $\tilde{\xi}_i(t) = \exp\left(-\left(\alpha_i\beta_i+\rho_i\right)t\right)\tilde{\xi}_i\left(0\right) + \alpha_i\int_0^t\left(\exp\left(-\left(\alpha_i\beta_i+\rho_i\right)\left(t-\tau\right)\right)\left|\zeta_i\left(\tau\right)\right|\right)\mathrm{d}\tau$. Since $\lim_{t\to\infty}\exp\left(-\left(\alpha_i\beta_i+\rho_i\right)\left(t-\tau\right)\right)\left|\zeta_i\left(\tau\right)\right|\mathrm{d}\tau = 0$, the integral $\int_0^t\exp\left(-\left(\alpha_i\beta_i+\rho_i\right)\left(t-\tau\right)\right)\left|\zeta_i\left(\tau\right)\right|\mathrm{d}\tau$ are UUB. Besides, $\lim_{t\to\infty}\exp\left(-\left(\alpha_i\beta_i+\rho_i\right)t\right)\tilde{\xi}_i\left(0\right) = 0$, then the UUB of $\tilde{\xi}_i(t)$ is proved. Let $b_{\tilde{\xi}_i}$ to be the upper bound of $\tilde{\xi}_i(t)$. Given Assumption 2, one has $\|\delta_i^{(p)}\| \leq \bar{\kappa}$, also based on (21), by choosing $\xi_i(0) > \bar{\kappa}-1$, and when $\left|\zeta_i(t)\right| \geq \max\left\{\frac{2\,\sigma_{\max}(\mathcal{L}+\mathcal{G})\Delta u_{fi}}{\sigma_{\min}(\mathcal{L}+\mathcal{G})}, \frac{p\alpha_i\beta_i b_{\tilde{\xi}_i}}{2m_1\bar{\kappa}^p}, \frac{\sigma_{\max}(\mathcal{L}+\mathcal{G})}{\sigma_{\min}(\mathcal{L}+\mathcal{G})}\right\}$, $\exists T$ such that $\forall t > T$, the sufficient condition (37) for $\dot{E} \leq 0$ is satisfied. Therefore, based on the LaSalle's invariance principle [10], $\zeta_i(t)$ is UUB. Since $(\mathcal{L}+\mathcal{G})$ is non-singular and $\zeta_i(t)$ is UUB, based on $\lim_{t\to\infty}\zeta = -\operatorname{diag}(c_i)(\mathcal{L}+\mathcal{G})\varepsilon$, ε is UUB, and proof completed. ■

979-8-3315-8138-1