

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2025.DOI

An AI-Enabled Hybrid Cyber-Physical Framework for Adaptive Control in Smart Grids

MUHAMMAD SIDDIQUE¹, SOHAIB ZAFAR²

¹Department of Electrical Engineering, NFC Institute of Engineering and Technology (NFC IET), Multan, Pakistan (e-mail: engr.siddique01@gmail.com)

²Lahore University of Management Sciences (LUMS), Lahore, Pakistan (e-mail: 17060003@lums.edu.pk)

Corresponding author: Muhammad Siddique (e-mail: engr.siddique01@gmail.com, msiddique@nfciet.edu.pk).

arXiv:2511.21590v1 [cs.LG] 26 Nov 2025

ABSTRACT Smart grids are a fusion of classical power infrastructure and advanced communication networks and smart control, to create a cyber-physical environment that is more efficient and flexible than ever before. This integration causes vulnerabilities that can undermine grid stability as well as reliability. Digital forensics is a fundamental concept of learning and identifying, detecting, and mitigating such security incidents. This paper presents an all-in-one machine learning-based digital forensic framework of smart grid systems deployed on the Cloud. The framework combines the data acquisition at the sensor-level, authenticated communication, scalable cloud storage and automated forensic analytics. The model uses supervised and unsupervised learning algorithms - such as Random Forest, Support Vector Machine, Gradient Boosted Trees and deep neural architectures for anomaly detection, event reconstruction and intrusion analysis in real time. After several simulation and experimental studies on real-time smart-meter data streams, the proposed framework is shown to be very accurate, scalable and resilient to cyber-attacks including data tampering, false-data injection and coordinated control-loop manipulation. The results indicate that cloud services are the best backbone for big-data-driven forensic workflows, which allows energy utilities to achieve a fast situational awareness and intelligent incident response.

INDEX TERMS Smart grid, digital forensics, machine learning, anomaly detection, cyber-physical systems, cloud computing, instrumentation analytics, Google Cloud Platform, .

LIST OF SYMBOLS

t	Continuous time variable	$C_G(P_G)$	Cost of conventional power generation
T	Final time horizon for optimization	$C_L(P_L)$	Cost of load curtailment
J	Objective cost functional	$C_E(E)$	Cost associated with energy storage
$u(t)$	Control input vector at time t (e.g., curtailment, charging actions)	$C_S(u)$	Cost of control action (e.g., switching, actuation)
$x(t)$	System state vector at time t (e.g., voltages, SOC)	η	Efficiency of energy conversion/storage
$y(t)$	Output vector (measured or controlled outputs)	λ	Lagrange multiplier (for constraints or dual formulations)
$A, B, C,$	System matrices for linear or linearized state-space model	δ	Perturbation or variation (e.g., in adaptive control)
D		τ	Time constant or control delay
P_G	Power generated by controllable sources	$P_{ij}(t)$	Power flow from bus i to bus j at time t
P_L	Power curtailed from load demand	Z_{ij}	Impedance of transmission line between nodes i and j
$R(t)$	Renewable generation at time t	$V_i(t)$	Voltage magnitude at bus i
$L(t)$	Total load demand at time t	$\theta_i(t)$	Voltage angle at bus i
E	Energy level in energy storage systems (e.g., batteries)	$f_i(t)$	Frequency at bus i
$S(t)$	State of charge (SOC) of energy storage	$\Delta f(t)$	Frequency deviation from nominal

$Q(x, u)$	State-action value function (in reinforcement learning)
r_t	Instantaneous reward at time t
γ	Discount factor for future rewards (RL)
π	Policy mapping state to actions in reinforcement learning
$\hat{x}(t)$	Estimated state (e.g., from observer or filter)
N	Prediction/control horizon in Model Predictive Control (MPC)
H	Hessian matrix in quadratic programming formulation
F	Forecast function for future renewables/load
σ	Standard deviation (used in uncertainty modeling)
\mathcal{P}	Set of prosumers or controllable participants
\mathcal{N}	Set of all network nodes (buses)
\mathcal{E}	Set of all network edges (lines)
B_{ij}	Susceptance between bus i and j
u_i^{\max}, u_i^{\min}	Upper and lower control bounds
E_{\max}, E_{\min}	Max/min energy limits for storage systems
P_G^{\max}	Maximum generation capacity of dispatchable generators
$D(t)$	Demand profile (possibly stochastic)
$W(t)$	Weather input (for forecasting solar/wind)
k	Discrete-time index (if applicable)

I. INTRODUCTION

Rising energy demand, urgent need for decarbonization, increasing penetration of stochastic renewable energy sources (RES), and electrification of transport have necessitated the transformation of the traditional electric grid to an intelligent, adaptive, and sustainable smart grid. [1], [2], [10]. Advancements in sensors have contributed to fusing the grid with a communication network and computational intelligence. The application of these advancements makes the monitoring, control, and optimization of the process possible in real-time across the layers. The aggregation of such a structure is what we know as a cyber-physical system(CPS). [11]–[13].

The system complexity and uncertainty at the operational level have significantly increased due to highly flexible loads (EVs) and integration of distributed renewable energy resources (DERs). [3], [14], [15], making traditional model-based control approaches ineffective [4], [16]. Particularly, such centralized architectures often face impediments due to latency, single-point failures, and scalability, reducing their effectiveness for the decentralized structure of the modern smart grids [5], [17]. Therefore, adaptable control architectures with standardized interfaces are warranted. They help resolve the core issue of effective coordination between spatio-temporally dispersed components, since these components may operate under different control regimes. The components may include DER units, microgrids, and EV Charging points. [18]–[20].

In addition to spatio-temporal decentrality, the monitoring and control solution for the modern grid needs to be

responsive to stochastic events. The monitoring and control solution must be responsive to highly stochastic events such as generation intermittency, load fluctuations, and device-level failures [21], [22]. Moreover, with the increasing fusion of information and communication technologies, the smart grid faces vulnerability vis-à-vis cyber attacks. Cyber attacks such as False Data Injection(FDI) and denial of service(DoS) have a demonstrated capability to undermine control systems and falsify state estimation [23]–[25]. Often, these attacks are stealthy and hence compromise the false data detection mechanisms in place, thereby risking the system stability. Lag introduced by the centralized cloud computing platforms is another central issue in the realization of a truly intelligent system. The network delays and resource allocation prove to be limiting the efficacy of otherwise scalable cloud systems. These limitations are especially a problem for time-sensitive grid operations such as voltage control, frequency regulation, and switch protection [26], [27]. Research has been done on edge and fog computing solutions for such problems. These solutions enable localized processing at the edge of the network, thereby improving the response time and resilience of the grid. [28], [29].

Apart from latency, interoperability is also of core concern. While industry standards have been developed to outline guidelines for interfacing and integrating heterogeneous systems [18], [19], [30], ensuring compliance throughout the legacy infrastructure and advanced technologies remains a continuing challenge.

Advancements in artificial intelligence (AI), especially in reinforcement learning and adaptive dynamic programming(ADP), offer a powerful data-driven approach towards finding the solution to complex decision-making problems in the context of increasing uncertainty. These “model-free” approaches have shown promise in direct learning from the environment [6], [7].

Furthermore,agent-based modeling (ABM) and game-theoretic paradigms present robust solutions to represent the behaviors of the grid participants, helping decentralized control and optimization within local and global constraints [8], [9]. However, incorporating AI-based methods within a harmonized hybrid CPS framework, which integrates both cyber and physical layers while ensuring adaptability and security, remains a research gap. Furthermore, limited work has been done on hybrid control schemes focused on combining localized edge intelligence with cloud-based optimization.

This paper proposes an *AI-enabled hybrid cyber–physical framework for adaptive control in smart grids*, seamlessly integrating agent-based modeling, reinforcement learning, and game theory using a layered architecture. The framework is based on two adaptive control methodologies: (1) a hybrid edge–cloud ADP controller, where local agents perform near-optimal control using local state estimations, while cloud-based agents refine global value functions using strategic coordination; and (2) a multi-agent deep reinforcement learning (DRL) scheme based on Proximal Policy Optimization (PPO) and Deep Q-Networks (DQN), enabling agents to learn from

interaction and adapt policies online in a model-free manner.

The architecture is augmented with a model for *cyber-physical resilience*. It includes FDI attack modeling and a resilience index, which quantifies the system's stability under extreme conditions. The modeling is contextualized in a hybrid CPS, mimicking the interaction and interplay of communication infrastructure, electrical devices, and user behavior and environmental uncertainty. This framework is validated through extensive simulations on the IEEE 33-Bus radial distribution system, connected with DERs, smart loads under multiple scenarios. The model is evaluated under normal operational conditions, load variability, and generation intermittency, as well as under cyber attacks through FDI. Compared to baseline models, the hybrid approach demonstrated better performance across metrics such as control cost, response time, system stability, and resilience index.

Contributions: The main contributions of this paper are summarized as follows:

- We propose a unified, hybrid CPS framework for smart grid control that integrates agent-based modeling, ADP, and DRL to enable adaptive and decentralized decision-making.
- A novel adaptive control algorithm is designed for a hybrid edge–cloud ADP controller, and a DRL-based scheme employing PPO and DQN for real-time control under uncertainty.
- A game-theoretic agent interaction model is formulated to enable strategic coordination among distributed agents while preserving individual autonomy.
- A cyber-resilience evaluation module is developed, incorporating FDI attack modeling and resilience index computation to assess system vulnerability and recovery capability.
- The effectiveness of the proposed framework is demonstrated through detailed simulation studies on a modified IEEE 33-bus test system under multiple operational and adversarial scenarios.

The rest of the paper is organized as follows: Section II presents the overview of the proposed hybrid CPS modeling framework. Section II-D describes the agent-based and game-theoretic modeling of the smart grid network. Section III details the design of the two adaptive control algorithms. Section II-E formulates the cyber–physical resilience metrics and FDI model. Section VII provides the simulation results and performance analysis. Finally, Section VIII concludes the paper with key insights and future directions.

II. SYSTEM OVERVIEW

This section describes the components constituting our model of the smart grid. A formal tuple is used to model the cyber-physical nature of the smart grid, as given below:

$$\mathcal{G}_{SG} = \langle \mathcal{P}, \mathcal{C}, \mathcal{E}, \mathcal{U}, \mathcal{O} \rangle$$

This abstract formulation mimics the character of the modern-day smart grid and enables a holistic analysis of its

principal components. The individual elements of the tuple are elaborated as follows:

- P: Physical power system** This consists of the components constituting the physical infrastructure, including generation, transmission, and distribution infrastructure behavior, load, and storage. These are modeled with differential algebraic equations governing their fundamental principles.
- C: Cyber-communication infrastructure** This component of the smart grid tuple models the communication protocols, smart meters, sensors, PMUs, and the SCADA system.
- E: Energy Management System (EMS)** The EMS is employed for generation and dispatch optimization as well as restoration operation of the grid.
- U: Control input space** Control input space includes generator setpoints, load shedding, DER control, load shedding and demand response.
- O: Objective space** The objective space is the solution space which includes the performance goals, including cost, reliability, emissions, and grid-resilience.

The tuple-based abstract modeling approach enables us to integrate AI-based adaptive control and reinforcement learning methodologies to ensure grid stability and resilience, efficient resource usage, and scalability of the model.

A. PHYSICAL LAYER

We comprehensively model the physical layer, consisting of the electrical infrastructure, including generation units, transmission, and distribution, as well as the network behavior. The physical layer is modeled for both steady-state and dynamic behaviors related to grid operation and control.

1) AC Power Flow

We model the power flow as the basic component of the physical layer using the AC power flow equations below:

$$\begin{aligned} P_i &= \sum_{j \in \mathcal{N}} |V_i||V_j|(G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}), \\ Q_i &= \sum_{j \in \mathcal{N}} |V_i||V_j|(G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}), \end{aligned} \quad (1)$$

Active power is quantitatively represented by (P_i) and reactive power injection at each bus is given by(Q_i). $|V_i|$ represent the voltage magnitudes and, θ_i represents angles respectively. Network admittance is given by G_{ij} , B_{ij} . The solution of the power flow equations provides the following:

- Voltage regulation and limit checks
- Optimal power dispatch
- Grid contingency analysis

AC power flow being included in the model is the foundation for larger optimization in the smart grid. Optimal Power Flow(OPF) and SCOPF stem from these basic equations.

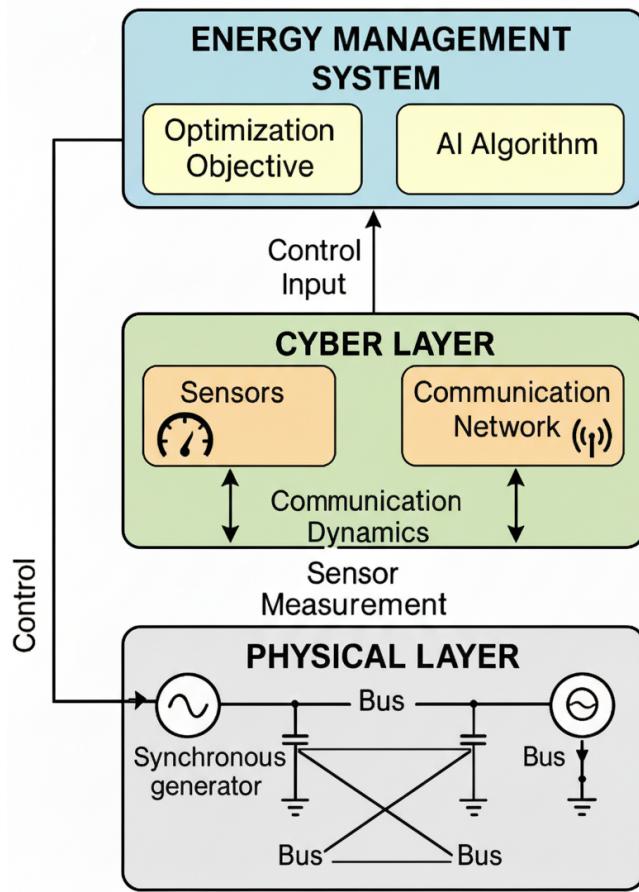


FIGURE 1: Architecture of hybrid cyber-physical modeling of smart grid for AI-based adaptive control.

2) Generator Dynamics

We model synchronous generator dynamics using the classical swing equation:

$$M_i \ddot{\delta}_i + D_i \dot{\delta}_i = P_{m,i} - P_{e,i} \quad (2)$$

where δ_i is the rotor angle of generator i , M_i is the inertia constant, D_i is the damping coefficient, and $P_{m,i}$ and $P_{e,i}$ denote the mechanical input and electrical output power, respectively.

a: Excitation System:

The excitation system involves the regulation of the generator's internal voltage $E'_{q,i}$ through the field voltage $E_{fd,i}$. The field voltage affects the electromotive force (EMF) of the generator, which is directly factored in the electrical power output, $P_{e,i}$ of the generator via the network. This interconnection of the variables couples the electrical and mechanical behavior of the machine. It is represented through a simplified model given by:

$$T_{A,i} \dot{E}_{fd,i} = -E_{fd,i} + K_{A,i} (V_{ref,i} - V_i) \quad (3)$$

where $T_{A,i}$ is the time constant of the automatic voltage regulator (AVR), $K_{A,i}$ is the AVR gain, $V_{ref,i}$ is the voltage

reference, and V_i is the terminal voltage magnitude of the generator.

b: Turbine-Governor System

A turbine governor is used to control the input mechanical power according to primary frequency control, which essentially bridges the mechanical dynamics to the frequency dynamics of the power system. The model for the turbine-governor system represents this relationship between mechanical power $P_{m,i}$ in response to frequency deviations. A first-order differential equation to model this is given as:

$$T_{g,i} \dot{P}_{m,i} = -P_{m,i} + P_{ref,i} - R_i^{-1} (\omega_i - \omega_0) \quad (4)$$

In this model, $T_{g,i}$ is the governor time constant, $P_{ref,i}$ is the mechanical power setpoint, R_i is the droop coefficient, $\omega_i = \dot{\delta}_i$ is the rotor speed, and ω_0 is the nominal system speed.

c: Algebraic Network Constraints

The nonlinear algebraic power flow equations are representative of the power balance at each bus and characterize all the generators and loads in a network. The electrical power output $P_{e,i}$ is computed from network conditions through these nonlinear algebraic power flow equations. These are given by:

$$P_{e,i} = \sum_{j \in \mathcal{N}} V_i V_j (G_{ij} \cos(\theta_i - \theta_j) + B_{ij} \sin(\theta_i - \theta_j)) \quad (5)$$

$$Q_{e,i} = \sum_{j \in \mathcal{N}} V_i V_j (G_{ij} \sin(\theta_i - \theta_j) - B_{ij} \cos(\theta_i - \theta_j)) \quad (6)$$

Here, V_i and θ_i are the voltage magnitude and phase angle at bus i , G_{ij} and B_{ij} are the conductance and susceptance components of the bus admittance matrix Y_{bus} , and \mathcal{N} is the set of all buses. These equations enforce power balance at each bus and couple the electrical behavior of all generators and loads in the network.

3) Full DAE System:

The complete system is described by a set of differential-algebraic equations (DAEs) of the form:

$$\dot{x} = f(x, y) \quad (7)$$

$$z = g(x, y) \quad (8)$$

In this formulation, x includes the dynamic state variables such as δ_i , ω_i , $E_{fd,i}$, and $P_{m,i}$, while y includes algebraic variables such as bus voltage magnitudes V_i and angles θ_i . The function $f(x,y)$ is a set of differential equations consisting of swing equation (2), excitation system (3) and turbine-governor model (4). The function $g(x, y)$ represents the algebraic constraints such as the power flow equations (5) and (6). This coupled DAE formulation is basic for simulating and analyzing the inter machine oscillations, dynamic stability and frequency response. It also provides the basis for

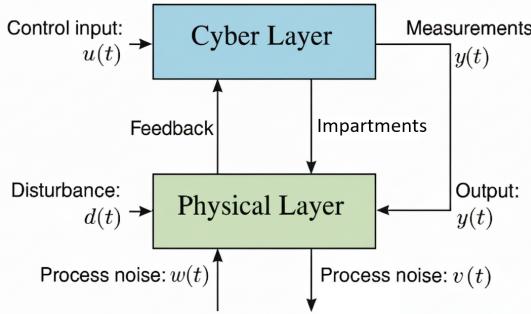


FIGURE 2: Illustration of the cyber-physical integration in a smart grid, showing interaction between the physical power system and the cyber infrastructure.

developing grid control strategies and integrating the renewable energy systems into traditional synchronous-machine-dominated power systems.

B. CYBER-COMMUNICATION INFRASTRUCTURE

This part is modeling the cyber communication infrastructure which is represented by \mathcal{C} in the tuple $\mathcal{G}_{SG} = \langle \mathcal{P}, \mathcal{C}, \mathcal{E}, \mathcal{U}, \mathcal{O} \rangle$. The cyber-communication infrastructure is comprised of the sensors, computation, communication, and information processing modules. In the interaction with the physical layer of the power system, the cyber-communication infrastructure ensures secure and trustworthy monitoring and control.

To elaborate, the cyber-communication layer includes technologies and components such as phasor measurement units (PMUs), smart meters, SCADA systems, intelligent electronic devices(IEDs), and edge computing infrastructure.

1) Sensor Measurement Models

The cyber-physical power system is highly influenced by the communication network, which interconnects cyber nodes such as substations, DER controllers, and EMS. This network allows for centralized operation as well as local autonomy. The dynamic states of the cyber nodes are affected by constraints introduced by the network limitations, such as packet losses, bandwidth constraints, etc.

To model the evolution of internal cyber-states under these conditions, we adopt a communication-aware nonlinear state-space formulation:

$$\begin{aligned}\dot{\hat{x}}_i(t) &= f_i(\hat{x}_i(t), u_i(t - \tau_{ij}(t)), d_i(t)) + B_{w_i} w_i(t), \\ y_i(t) &= h_i(\hat{x}_i(t)) + v_i(t),\end{aligned}\quad (9)$$

Here, the $\hat{x}_i(t) \in \mathbb{R}^{n_i}$ represents the internal cyber-state vector of node i . Node includes the relevant digital variables relevant to the local control logic. Further, the term $u_i(t - \tau_{ij}(t)) \in \mathbb{R}^{m_i}$ mimics the control input from node j . This control input is affected by time-varying communication lag, represented in our formulation as $\tau_{ij}(t)$, capturing latency induced due to congestion or asynchronous

scheduling. Moreover, $d_i(t) \in \mathbb{R}^{p_i}$ mimics the exogenous disturbances such as protocol jitter, system noise, and time drift at i . Next, $w_i(t)$ models communication-induced uncertainties, such as jitter, variable latency, or cyber attacks. $v_i(t)$ represents measurement noise and digital corruption in sensor and control signals. The matrix B_{w_i} governs how such disturbances propagate into the cyber-state dynamics of node i . Lastly, output vector $y_i(t) \in \mathbb{R}^{q_i}$, includes raw or processed data streams, timestamps, delay acknowledgements or other observable communication metrics at node i .

Furthermore, $f_i(\cdot)$ models the evolution of control law implementation, local optimization routines, and protocol stack processing i . The function $h_i(\cdot)$ mimics the output logic of the node. This includes sensing feedback, diagnostic responses, and actuation signals.

The communication-induced phenomena of the modern smart-grid environment are given by a comprehensive framework represented in Equation (9). Communication latency and jitter are represented by $\tau_{ij}(t)$. $d_i(t)$ and $w_i(t)$ model the exogenous disturbance inputs and stochastic noise components, respectively. In addition, the scheduling delays and protocol-induced errors are embedded within the non-linear structure of the function $f_i(\cdot)$. The clock drift and synchronization mismatches are implicitly modeled within the internal state formulation, allowing the model to represent asynchronous interactions and temporal misalignment across the system.

C. ENERGY MANAGEMENT SYSTEM

EMS is the central cyber-physical component in charge of real-time optimization. It is responsible for coordination of distributed generation, controllable loads and storage units. It receives the data from sensors, it forecasts and then has the ability to execute control logic and communicate the control decisions to the field devices and operators. Altogether, it is the brain of the system striving to achieve system's operational efficiency, reliability and sustainability. This section elaborates the modeling of the EMS.

1) Optimization Objective

The first optimization task of the EMS is to ascertain the control inputs $u(t)$ to minimize the overall operational cost of the system in a given time horizon $[0, T]$.

$$\min_{u(t)} J = \int_0^T \left(C_G(P_G) + C_L(P_L) + C_E(E) + C_S(u) \right) dt. \quad (10)$$

Here:

- $C_G(P_G)$: The cost of the generation based on the generation unit's fuel and ramp rates.
- $C_L(P_L)$: Load Curtailment penalty which essentially reflects customer response.
- $C_E(E)$: Cost of energy storage. This abstracts degradation of battery life and associated dispatch penalties.

- $C_S(u)$: This abstracts cost of control. This is employed to discourage frequent control actions, which induce instability in the system.

The EMS objective combines overall performance over a period, while ensuring multi-objective goals of low operational cost, low emissions, and control optimization are met, keeping the maximization of service reliability as a central piece.

Constraints: The constraints on the objective function described above are grounded in the first-principles involved in the power system. The are given as :

$$\begin{aligned} f_{PF}(x, u) &= 0, \\ P_{G_i}^{\min} &\leq P_{G_i}(t) \leq P_{G_i}^{\max}, \\ \dot{E}_i(t) &= \eta_{ch} P_{ch,i}(t) - \frac{1}{\eta_{dis}} P_{dis,i}(t), \\ 0 &\leq P_{shed,i}(t) \leq P_{L_i}(t). \end{aligned} \quad (11)$$

Where:

- $f_{PF}(x, u) = 0$: This is a nonlinear equality constraint. It enforces the physical feasibility of voltage and current variables.
- $P_{G_i}^{\min}, P_{G_i}^{\max}$: System's Minimum and Maximum capacity of the respective generators. This reflects technical and economic dispatch boundaries.
- $\dot{E}_i(t)$: Reflects the rate of storage for unit i .
- η_{ch}, η_{dis} : Charging and discharging efficiency factors (typically $0.85 \leq \eta \leq 0.98$) modeling converter/inverter losses.
- $P_{ch,i}(t), P_{dis,i}(t)$: At any given time t , the charging and discharging power of the storage units i .
- $P_{shed,i}(t)$: Load shedding by a given bus i . This is limited to not cross the demand at any given time $P_{L_i}(t)$.

The constraints enable security and sustainability of the operation alongside the integration of the storage, renewables, and flexible demands.

D. AGENT BASED AND GAME THEORETIC MODELING

Unlike the traditional grid, modern smart grids are a decentralized and dynamic network of disparate agents, including domestic units, EVs, DER, and energy storage units. These agents interact with the grid bidirectionally and are often called *prosumers*, which produce as well as consume the energy. This highly dynamic behavior can be modeled and optimized through advanced agent-based modeling methodologies and game-theoretic frameworks. These methods present an effective method for distributed decision making suitable for decentralized systems such as the smart grid.

Utilizing this, we present an agent-based prosumers optimization method in this section. We also present a non-cooperative game-theoretic framework mimicking the strategic interactions between agents for various available resources.

1) Prosumer Optimization

Each prosumer $i \in \mathcal{P}$ aims to minimize its individual cost function over a planning horizon. The cost function typically accounts for electricity consumption, local generation (e.g., solar PV), battery usage, and potential monetary incentives or penalties. A basic optimization problem for a prosumer is given by:

$$\mathcal{A}_i = \arg \min_{P_i(t)} [c_i(P_i(t)) - \lambda(t)P_i(t)], \quad (12)$$

where:

- $P_i(t)$: Net power exported (positive) or imported (negative) by prosumer i at time t .
- $c_i(P_i(t))$: Cost or disutility function, typically convex (e.g., quadratic) representing discomfort, battery degradation, or fuel usage.
- $\lambda(t)$: Dynamic locational marginal price (LMP) or incentive rate broadcast by the system operator.
- \mathcal{A}_i : The action or strategy space of agent i , such as consumption level, charging schedule, or generation output.

The optimization can be subject to local constraints:

$$0 \leq P_i^{\text{load}}(t) \leq P_i^{\text{load,max}}, \quad (13)$$

$$0 \leq P_i^{\text{gen}}(t) \leq P_i^{\text{gen,max}}, \quad (14)$$

$$0 \leq S_i(t) \leq S_i^{\max}, \quad (15)$$

where $S_i(t)$ is the state of charge of the local battery.

More sophisticated versions include forecast-based optimization using Model Predictive Control (MPC), where each agent solves:

$$\min_{P_i(t:t+N)} \sum_{k=t}^{t+N} [c_i(P_i(k)) - \lambda(k)P_i(k)], \quad (16)$$

for a prediction horizon N . This supports dynamic decision-making under uncertainty in demand or PV output.

2) Non-Cooperative Game

In environments where multiple prosumers act simultaneously and selfishly, their actions impact each other through shared variables like voltage, frequency, or market prices. This leads to a non-cooperative game setup:

$$\max_{\mathcal{A}_i} U_i(\mathcal{A}_i, \mathcal{A}_{-i}), \quad \forall i \in \mathcal{N}, \quad (17)$$

where:

- $U_i(\cdot)$: Utility function of agent i .
- \mathcal{A}_i : Strategy of agent i .
- \mathcal{A}_{-i} : Strategies of all other agents ($j \neq i$).
- \mathcal{N} : Set of all participating agents or prosumers.

A classical solution concept is the **Nash Equilibrium**, where no agent can unilaterally improve its utility:

$$U_i(\mathcal{A}_i^*, \mathcal{A}_{-i}^*) \geq U_i(\mathcal{A}_i, \mathcal{A}_{-i}^*), \quad \forall \mathcal{A}_i \in \mathcal{A}_i, \forall i.$$

Utility functions U_i can incorporate:

- Energy profit: $\lambda(t)P_i(t)$,
- Battery wear cost: $\beta \cdot \text{DoD}_i(t)$,
- Comfort or preference: deviation from desired HVAC or EV charging profile,
- Peer-to-peer (P2P) trade benefits or penalties.

The existence of a Nash equilibrium is guaranteed under certain conditions such as convexity of utility functions and compactness of strategy sets. For uniqueness, strict convexity or contraction mappings may be needed. Game-theoretic analysis often leads to distributed algorithms, e.g., best-response dynamics or fictitious play.

E. SECURITY AND RESILIENCE

As the IoT devices and communication infrastructure is integrated in smart grids, it becomes imperative to ensure strong cybersecurity measures. This section explains the False Data Injections and assessment of system resilience.

1) False Data Injection

One of the most prominent and dangerous classes of cyberattacks in smart grids is the False Data Injection (FDI) attack. FDI attacks aim to compromise the integrity of sensor measurements by inserting malicious data that can mislead state estimation, control actions, or optimization routines without being detected by standard bad-data detection schemes. Mathematically, if $y_i(t)$ is the true sensor measurement at time t , then the compromised measurement under an FDI attack is modeled as:

$$\tilde{y}_i(t) = y_i(t) + a_i(t), \quad (18)$$

where $a_i(t)$ denotes the adversarial signal injected into the measurement. In a coordinated attack, the adversary may exploit the topology of the communication graph and knowledge of the system model to design $a_i(t)$ such that the corrupted measurement passes standard residual checks used in state estimation algorithms. Advanced techniques for FDI detection include machine learning-based anomaly detection, robust state estimation, and statistical change detection. However, securing all communication nodes and sensors in large-scale smart grids remains a nontrivial challenge due to cost and scalability concerns.

2) Resilience Index

System resilience defines the grid's ability to perceive, absorb, adapt and recover from disruptions. The disruptions may be cyberattacks but also physical faults and natural disasters such as wildfires. Resilience Index(RI) is the metric which comparatively quantifies the system's deviation from its nominal behavior during a disturbances.

The resilience index used for evaluation is

$$R = 1 - \frac{\sum_k \|x_k - x_k^{\text{nom}}\|^2}{\sum_k \|x_k^{\text{nom}}\|^2}, \quad (19)$$

where x_k^{nom} is the nominal state trajectory. This metric, computed directly in the simulation, quantifies the ability

of the controller to maintain stable, nominal-like operation during disturbances and cyberattacks. Factors affecting the resilience include deployment of redundant and diverse control mechanisms, EMS responsiveness to anomalous behavior, the strength of the communication network, and the autonomous ability of AI-based adaptive controllers to reconfigure operations.

III. ADAPTIVE CONTROL METHODOLOGY

The adaptive control layer governs the control response to external disturbance and prosumers' behavior. The adaptive control is scalable and robust due to its ability to learn directly from data and not the models. Two foundations in our approach are adaptive dynamic programming(ADP), Proximal Policy Optimization (PPO) and Deep Q-Network (DQN) for long-term performance.

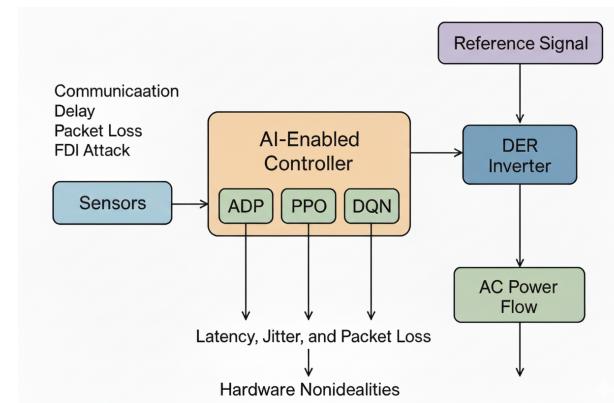


FIGURE 3: Hybrid AI-enabled cyber-physical control architecture illustrating sensing, communication delay, packet loss, AI controllers (ADP, PPO, DQN), inverter actuation, and AC power flow dynamics.

A. ADAPTIVE DYNAMIC PROGRAMMING

Adaptive Dynamic Programming (ADP) provides a framework to approximate solutions to complex dynamic optimization problems using learning-based techniques. It extends Bellman's principle of optimality to problems with unknown or nonlinear dynamics.

The core of ADP involves minimizing the expected cumulative cost-to-go. For a state x_t , control u_t , and cost function $r(x, u)$, the optimal control $u^*(t)$ satisfies:

$$u^*(t) = \arg \min_u \left[r(x_t, u_t) + \gamma \hat{V}(x_{t+1}) \right], \quad (20)$$

where $r(x_t, u_t)$ is the immediate cost incurred at time t , $\hat{V}(x_{t+1})$ is the learned value function that approximates the expected future cost starting from state x_{t+1} , and $\gamma \in (0, 1]$ is the discount factor that reflects the relative importance of future rewards. The learned value function $\hat{V}(x)$ serves as an

approximation to the true optimal cost-to-go function $V^*(x)$, which is defined as:

$$V^*(x_t) = \min_{\{u_k\}_{k=t}^{\infty}} \sum_{k=t}^{\infty} \gamma^{k-t} r(x_k, u_k), \quad (21)$$

subject to the system dynamics $x_{k+1} = f(x_k, u_k)$. Since the exact function $V^*(x)$ is generally unknown or intractable in high-dimensional, nonlinear systems, ADP approximates it using a parameterized function:

$$\hat{V}(x; \theta) = \sum_{i=1}^N \theta_i \phi_i(x) = \Phi(x)^\top \theta, \quad (22)$$

where $\phi_i(x)$ are chosen basis functions (e.g., radial basis functions, polynomials, or neural network activations), θ_i are trainable weights, and $\Phi(x)$ is the feature vector. Alternatively, $\hat{V}(x)$ may be implemented using a neural network denoted by $\mathcal{N}_\theta(x)$, where θ represents the learnable parameters of the network.

The immediate cost function $r(x_k, u_k)$ quantifies the penalty associated with the current state and control input. It is typically designed to penalize deviations from desired operating conditions and to discourage excessive control effort. A commonly used quadratic form is:

$$r(x_k, u_k) = (x_k - x_{\text{ref}})^\top Q (x_k - x_{\text{ref}}) + u_k^\top R u_k, \quad (23)$$

where x_{ref} is the reference state (e.g., nominal frequency or voltage), $Q \succeq 0$ is a state weighting matrix, and $R \succ 0$ is a control weighting matrix. This cost function penalizes deviations from the states as well as the control actions so it ensures stable and efficient operation. A simple form of the cost function has the form of:

$$r(x_k, u_k) = \omega_k^2 + \alpha u_k^2, \quad (24)$$

where ω_k is the frequency deviation and u_k is the control input from the governor or AVR, and $\alpha > 0$ controls the trade-off between tracking performance and control effort. By iteratively updating both the value function approximation $\hat{V}(x)$ and the optimal control $u^*(t)$, ADP enables near-optimal decision-making in real time, even in the presence of system uncertainties and nonlinear dynamics. In smart grids, this framework allows agents (e.g., DER controllers, load-serving entities) to update control actions in real-time based on observed system evolution. Unlike traditional dynamic programming, ADP handles high-dimensional state-action spaces via function approximators (e.g., neural networks).

B. PROXIMAL POLICY OPTIMIZATION (PPO) CONTROL METHODOLOGY

Proximal Policy Optimization (PPO) is a policy-gradient reinforcement learning algorithm that enables stable and sample-efficient control in nonlinear dynamic systems. It is particularly suitable for smart grid applications where the control space is continuous (e.g., inverter reactive power modulation, EV charging rates) and where measurement disturbances, cyber-latency, and stochastic behavior of loads

require robustness. PPO optimizes a parameterized control policy $\pi_\theta(u|x)$ that maps system states to control actions, enabling fast real-time decision-making in decentralized energy resources.

The goal of PPO is to maximize the expected return:

$$J(\theta) = \mathbb{E}_{\pi_\theta} \left[\sum_{k=t}^{\infty} \gamma^{k-t} r(x_k, u_k) \right], \quad (25)$$

where $\gamma \in (0, 1]$ is the discount factor and $r(x_k, u_k)$ is the instantaneous cost or reward derived from system performance. In the context of frequency and voltage regulation in a grid, the reward typically penalizes frequency deviations, voltage violations, and excessive inverter effort. Unlike classical policy gradient methods, PPO constrains policy updates using a clipped surrogate objective designed to prevent large detrimental gradient steps. The clipped objective for PPO is:

$$L^{\text{clip}}(\theta) = \mathbb{E}_t \left[\min \left(r_t(\theta) \hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \right], \quad (26)$$

where

$$r_t(\theta) = \frac{\pi_\theta(u_t | x_t)}{\pi_{\theta_{\text{old}}}(u_t | x_t)}$$

is the likelihood ratio and \hat{A}_t is the advantage estimate computed using generalized advantage estimation (GAE). The clipping threshold ϵ limits how far the new policy can deviate from the old one at each update, significantly improving stability under noisy smart-grid measurements.

The agent uses a parameterized policy network (actor) to output continuous control actions:

$$u_t = \pi_\theta(x_t),$$

and a value network (critic) to approximate the state-value function:

$$V_\psi(x_t) \approx \hat{V}(x_t) = \mathbb{E} \left[\sum_{k=t}^{\infty} \gamma^{k-t} r(x_k, u_k) \right].$$

In smart grid applications, the state vector includes bus voltages, angles, renewable power availability, SOC of batteries, EV charging levels, and cyber-layer indicators such as delay and packet loss. PPO updates both the actor and critic networks with batched trajectory data from the cyber – physical simulation environment. This leads to a robust controller that is able to compensate the renewable intermittency and measure corruption and stabilize system frequency.

To specialize PPO to perform grid control, the reward function is defined as follows:

$$r(x_k, u_k) = -(\omega_k^2 + \beta_v(V_k - 1)^2 + \alpha \|u_k\|^2), \quad (27)$$

where ω_k is the frequency deviation, V_k is the voltage magnitude and $\alpha, \beta_v > 0$ are the penalties for excessive control effort and voltage deviations. This reward structure enables PPO to learn smooth and stable, safe control actions even in the presence of network delay, cyber-attacks, and hardware uncertainties.

In the proposed hybrid control framework, PPO is the high-level continuous-action controller that is in charge of real-time tuning of inverter outputs, battery charge/discharge levels and flexible load-tuning. To noisy gradients and its capacity to manage huge nonlinear state spaces make it suitable for next generation adaptive smart grid operations.

C. DEEP Q-NETWORK (DQN) CONTROL METHODOLOGY

Deep Q-Networks (DQN) is a value-based reinforcement learning algorithm that is applicable for discrete control problems in cyber-physical smart grids. While PPO is used for continuous control signals, DQN is used for discrete decision-making like mode switching (charging/discharging states of storage,) EV charging priority assignment, load-shedding triggers and inverter operation modes. The method is an extension to classical Q-learning in which $Q(x, u)$ is approximated by a deep neural network which allows efficient learning in large nonlinear state spaces.

DQN aims to approximate the optimal Q -function which is given by the Bellman optimality equation:

$$Q^*(x_t, u_t) = r(x_t, u_t) + \gamma \max_{u'} Q^*(x_{t+1}, u'), \quad (28)$$

which represents the maximum future return achievable from state x_t by taking action u_t . Since directly computing Q^* is intractable for high-dimensional systems like smart grids, DQN uses a parameterized function $Q_\theta(x, u)$ to approximate it:

$$Q_\theta(x_t, u_t) \approx Q^*(x_t, u_t),$$

where θ denotes the neural network parameters. The network parameters are updated by minimizing the temporal-difference loss:

$$\mathcal{L}(\theta) = \left(r_t + \gamma \max_{u'} Q_{\theta^-}(x_{t+1}, u') - Q_\theta(x_t, u_t) \right)^2, \quad (29)$$

where Q_{θ^-} is the target network, updated slowly to improve training stability. Experience replay buffers further decorrelate training samples, preventing divergence and enabling stable learning even when the environment is highly nonlinear and stochastic. In the context of the proposed smart grid architecture, the state x_t contains voltage profiles, renewable forecasts, load levels, SOC of batteries, and cyber-health indicators (latency, packet success rate). The discrete action set includes:

- battery mode selection (charge/discharge/idle),
- EV charging priority states,
- inverter tap or mode changes,
- load-shedding or restoration triggers.

To tailor the DQN controller for grid stability, the reward function is defined as:

$$r(x_k, u_k) = - (\omega_k^2 + \lambda \mathbb{I}\{\text{voltage violation}\} + \alpha \|u_k\|^2), \quad (30)$$

where $\mathbb{I}\{\cdot\}$ denotes an indicator function and λ penalizes actions that result in unacceptable voltage profiles. This reward

formulation encourages DQN to select discrete operational modes that indirectly support frequency and voltage stability, especially during fast renewable ramps or cyber-induced disturbances. DQN is thus integrated into the hybrid control framework as a complementary discrete-action controller operating alongside the continuous-action PPO controller. Together with the ADP layer, DQN contributes to hierarchical decision making where discrete grid management actions are coordinated with continuous inverter and storage control signals, creating a unified adaptive control mechanism for complex cyber-physical smart grids.

IV. MODELING AND IMPLEMENTATION OF TESTBED

The physical power system, measurement infrastructure, and cyber-physical interaction architecture on which the proposed control methodology is implemented emulated in the simulation environment which consist of a physical layer i.e. modified IEEE 33-bus radial distribution system, enriched with distributed energy resources (DERs), electric vehicles (EVs), responsive loads, and supervisory communication layers. The physical layer is tightly coupled with the cyber layer, enabling realistic evaluation under communication delay, packet drops, measurement corruption, and cyberattacks.

The network is modeled using a nonlinear differential-algebraic system:

$$\dot{x} = f(x, y, u), \quad (31)$$

$$z = g(x, y), \quad (32)$$

where x represents dynamic states (e.g., inverter states, SOC), y contains algebraic variables (voltages, angles), and u represents the control actions synthesized by the hybrid RL-based controller. The physical power flows are computed using full nonlinear AC equations, ensuring realism in voltage, frequency, and power exchange behavior.

A. LOADS AND DEMAND PROFILES

The loads at each bus are modeled as time-varying active and reactive power demands:

$$P_{L,k} = P_L^{\text{base}}(1 + \Delta P_k), \quad (33)$$

$$Q_{L,k} = Q_L^{\text{base}}(1 + \Delta Q_k), \quad (34)$$

where the perturbations ΔP_k and ΔQ_k follow stochastic time-series patterns emulating residential, commercial, and EV-driven peak variations. Sudden changes of up to 20% are introduced to emulate high-variability events, forcing the controller to react in real time. An aggregated EV charging block is modeled as an adjustable active power load E_k^{EV} , affecting both voltage stability and system frequency during high charging demand.

B. RENEWABLE GENERATION AND DER MODELING

The network includes solar PV and wind DER units placed at selected buses. Their injections follow:

$$S_k^{\text{solar}} = P_{\text{PV}}(k) + jQ_{\text{PV}}(k), \quad (35)$$

$$S_k^{\text{wind}} = P_{\text{W}}(k) + jQ_{\text{W}}(k), \quad (36)$$

with their active power governed by real irradiance and wind-speed profiles:

$$P_{\text{PV}}(k) = \eta_{\text{PV}} A_{\text{PV}} I_{\text{solar}}(k), \quad (37)$$

$$P_{\text{W}}(k) = \frac{1}{2} \rho A C_p v_{\text{wind}}^3(k). \quad (38)$$

To emulate renewable intermittency, abrupt ramps and cloud-induced drops of 20–40% are introduced. These disturbances strongly impact bus voltages and frequency, making them ideal for testing adaptive RL controllers.

C. BATTERY STORAGE AND EV MODELING

Battery storage units are included with the state-of-charge (SOC) dynamic:

$$SOC_{k+1} = SOC_k + \eta_c P_k^{\text{ch}} - \frac{1}{\eta_d} P_k^{\text{dis}}, \quad (39)$$

where P^{ch} and P^{dis} represent charging/discharging commands derived from the selected controller. SOC affects both frequency regulation and voltage stability through local balancing. EVs are modeled as controlled loads whose demand E_k^{EV} can be modulated by the RL agents to maintain grid stability during congested periods.

D. MEASUREMENT LAYER AND STATE ACQUISITION

Each bus is assumed to have a PMU-like sensing layer capable of reporting:

$$s_k = [V_k, \theta_k, P_{L,k}, Q_{L,k}, S_k^{\text{solar}}, S_k^{\text{wind}}, SOC_k, E_k^{\text{EV}}, f_k]^\top \quad (40)$$

These measurements form the state vector used by all controllers. Sensor update rates are 1 second, matching real-time distribution-level monitoring systems. Measurement noise is modeled as:

$$v_k \sim \mathcal{N}(0, \sigma_v^2), \quad (41)$$

with σ_v^2 chosen to represent typical PMU and micro-PMU noise characteristics.

E. CYBER LAYER, COMMUNICATION MODEL, AND ATTACKS

The communication layer is modeled using a time-varying latency variable τ_k (0–250 ms), a packet-loss probability p_{drop} , and an FDI attack vector a_k :

$$\tilde{s}_k = \begin{cases} s_{k-\tau_k} + v_k + a_k, & \text{with probability } 1 - p_{\text{drop}}, \\ s_{k-1}, & \text{if a packet is dropped.} \end{cases} \quad (42)$$

The FDI attacks manipulate voltage magnitude and frequency to mislead controllers. Time-synchronized bursts occur randomly to emulate coordinated cyberattacks. This cyber-physical coupling makes it necessary for the controller to constantly cope with delayed or outdated states, with corrupted measurements, with missing communications, and even on instability caused by an attack.

F. ACTUATION AND INVERTER CONTROL MODEL

Controller outputs affect the physical network through inverter-based DERs and controllable loads. The applied action includes edge noise:

$$u_k^{\text{applied}} = u_k + e_k, \quad e_k \sim \mathcal{N}(0, \sigma_{\text{edge}}^2), \quad (43)$$

representing inverter switching nonidealities and real hardware uncertainty. Typical actions of inverters are reactive power injection or absorption, active power curtailment, voltage reference modification and frequency support. These actions influence the subsequent AC power-flow solution and thus the next state s_{k+1} . The full system is comprised of a nonlinear radial distribution grid and time-varying loads and renewable energy sources, battery and EV dynamics, PMU-like sensing devices, communication delays and noise, cyberattacks in the form of false data injection (FDI), and various actuator noise sources and inverter nonidealities.

This tightly integrated cyber-physical structure provides a realistic platform for deploying the hybrid ADP-PPO-DQN control methodology described in Section VI. The system reflects the operational conditions of next-generation smart grids, where both physical disturbances and cyber vulnerabilities must be simultaneously addressed.

V. PRACTICAL PROTOTYPIC SYSTEM SPECIFICATIONS

To validate the hybrid cyber-physical control methodology described in Section III, a detailed prototypic distribution-level power system was constructed in the simulation environment. The system emulates a medium-voltage radial feeder equipped with distributed energy resources (DERs), electric vehicles (EVs), and advanced metering infrastructure (AMI). The numerical values of system parameters, communication characteristics, and physical component models were selected to reflect realistic operating conditions typically observed in practical distribution networks such as the IEEE 33-bus.

The following subsections summarize the physical and cyber parameters used for the implementation of the proposed control methodology.

A. NETWORK TOPOLOGY AND ELECTRICAL PARAMETERS

The prototypic grid consists of a radial feeder with 33 buses, a base voltage of 12.66 kV, a base power of 10 MVA, and a nominal frequency of 50 Hz. The feeder contains a mix of residential, commercial, and small industrial loads with time-varying characteristics. Transformer parameters are modeled with a resistance of $R_{\text{tr}} = 0.01$ p.u., a reactance of $X_{\text{tr}} = 0.04$ p.u., and a load-to-no-load ratio of $T_{\text{load}}/T_{\text{no-load}} = 0.8/0.2$. For line segments, typical underground or overhead conductor impedances are assumed, with resistance $R_\ell \in [0.03, 0.09]$ Ω/km and reactance $X_\ell \in [0.04, 0.12]$ Ω/km, and segment lengths ranging from 0.6 km to 1.9 km depending on the bus number. Nodal voltage limits are enforced according to standard utility practice, with $0.95 \leq V_k \leq 1.05$ p.u..

B. LOAD, RENEWABLE, AND STORAGE MODELING

The physical system incorporates time-varying loads, stochastic renewable generation, and distributed storage units to emulate real-world operational variability. The load profiles include active loads $P_{L,k} \in [0.2, 1.8]$ MW, reactive loads $Q_{L,k} \in [0.05, 0.9]$ MVar, daily variability of 15–30%, and stochastic fluctuations modeled as Gaussian noise with $\sigma_P = 0.03$ MW. The solar PV system has a rated power of $S_{k,solar,max} = 500$ kW, with intermittency modeled via Beta-distributed irradiance and cloud-induced dips ranging from 20–60% of nominal output. The wind generator features a rated power of $S_{k,wind,max} = 300$ kW, a wind speed following a Weibull distribution with scale $c = 7.5$, and cut-in/cut-out limits between 3–22 m/s. The battery energy storage system (BESS) is modeled with a maximum energy capacity of $E_{max} = 300$ kWh, maximum charge/discharge power $P_{max} = \pm 150$ kW, and state-of-charge limits $SOC_k \in [0.20, 0.90]$, with a charging/discharging efficiency of 95%. The electric vehicle charging load ranges from $E_k^{EV} \in [40, 110]$ kW with an efficiency of $\eta_{EV} = 0.92$.

C. CYBER LAYER, NETWORK DELAY, AND ATTACK PARAMETERS

The communication layer emulates practical conditions in AMI and edge-to-cloud coordination, where each state vector reported by bus controllers is affected by latency, packet drops, measurement noise, and false data injection (FDI) attacks. The communication delay is modeled as a discrete-time variable $\tau_k \in \{0, 1, 2, 3\}$ steps (equivalent to 20–120 ms), following a truncated Gaussian distribution. Packet loss is represented by a probability $p_{drop} = 0.05$, reflecting congestion and wireless interference typical in edge networks. Measurement noise is included for voltage and frequency, with $v_{V,k} \sim \mathcal{N}(0, 0.005^2)$ and $v_{f,k} \sim \mathcal{N}(0, 0.02^2)$, consistent with real PMU/μPMU accuracy. FDI attacks modify the voltage and frequency as $a_{V,k} \in [-0.03, 0.03]$ p.u. and $a_{f,k} \in [-0.15, 0.18]$ Hz, applied with a probability $p_{FDI} = 0.04$ during attack intervals. Control signals sent to DER inverters experience physical uncertainty:

$$u_k^{\text{applied}} = u_k + e_k, \quad e_k \sim \mathcal{N}(0, 0.01^2). \quad (44)$$

Inverter response dynamics follow a first-order model:

$$\tau_{inv} \dot{P} + P = P_{ref}, \quad \tau_{inv} = 40 \text{ ms}. \quad (45)$$

VI. IMPLEMENTATION OF CONTROL METHODOLOGY

This section presents the mathematical formulation of the hybrid cyber-physical control architecture implemented in the simulation engine. The controller integrates Adaptive Dynamic Programming (ADP), Proximal Policy Optimization (PPO), and Deep Q-Networks (DQN) under communication constraints, packet drops, measurement corruption, and false-data injection (FDI) cyberattacks.

A. STATE VECTOR AND CYBER OBSERVATION MODEL

At each discrete control step k , each bus provides a measurement vector

$$s_k = [V_k, \theta_k, P_{L,k}, Q_{L,k}, S_k^{\text{solar}}, S_k^{\text{wind}}, SOC_k, E_k^{EV}, f_k]^T \quad (46)$$

where V_k is voltage (p.u.), θ_k is angle (rad), $P_{L,k}$ and $Q_{L,k}$ are active/reactive loads, S_k^{solar} and S_k^{wind} are renewable injections, SOC_k is battery state-of-charge, E_k^{EV} is EV charging load, and f_k is frequency (Hz). These correspond directly to the features normalized in the simulation. The cyber layer distorts the measurement due to delay, packet loss, noise, and FDI attacks. Let τ_k be the communication delay, p_{drop} the packet loss probability, v_k the measurement noise, and a_k the FDI corruption. The controller receives

$$\tilde{s}_k = \begin{cases} s_{k-\tau_k} + a_k + v_k, & \text{with probability } 1 - p_{drop}, \\ s_{k-1}, & \text{if a packet drop occurs.} \end{cases} \quad (47)$$

here delays are discretized, FDI modifies voltage and frequency, and packet loss forces fallback to the previous state.

B. UNIFIED CONTROL OBJECTIVE

All controllers minimize the same quadratic cost

$$c_k = (V_k - 1)^2 + (f_k - 50)^2 + \alpha \sum_{i=1}^m u_{k,i}^2, \quad (48)$$

which penalizes deviations from nominal voltage and frequency and includes a control-effort penalty weighted by $\alpha > 0$. The reward used in reinforcement learning is

$$r_k = -c_k. \quad (49)$$

C. ADAPTIVE DYNAMIC PROGRAMMING (ADP)

ADP approximates the value function

$$V(s_k) \approx \hat{V}(s_k; \phi), \quad (50)$$

where ϕ denotes neural network parameters. The policy network produces a continuous control action

$$u_k^{\text{ADP}} = \pi(s_k; \theta), \quad (51)$$

with parameters θ . The temporal-difference (TD) target and error are

$$y_k = r_k + \gamma \hat{V}(s_{k+1}; \phi), \quad \delta_k = y_k - \hat{V}(s_k; \phi). \quad (52)$$

Both edge and cloud ADP controllers are implemented:

$$u_k^{\text{edge}} = \pi(\tilde{s}_k), \quad (53)$$

$$u_k^{\text{cloud}} = \pi(s_k), \quad (54)$$

where the cloud version benefits from cleaner states but suffers communication delay.

D. PROXIMAL POLICY OPTIMIZATION (PPO)

PPO uses a Gaussian stochastic policy

$$u_k^{\text{PPO}} \sim \mathcal{N}(\mu_\theta(\tilde{s}_k), \sigma_\theta^2(\tilde{s}_k)), \quad (55)$$

where μ_θ and σ_θ are outputs of the actor network. The importance ratio is

$$\rho_k(\theta) = \frac{\pi_\theta(a_k | s_k)}{\pi_{\theta_{\text{old}}}(a_k | s_k)}, \quad (56)$$

and the clipped PPO objective is

$$L^{\text{PPO}}(\theta) = \mathbb{E}[\min(\rho_k(\theta)A_k, \text{clip}(\rho_k(\theta), 1 - \epsilon, 1 + \epsilon)A_k)], \quad (57)$$

with advantage estimate $A_k = r_k - V(s_k)$. Packet loss is implemented as

$$u_k^{\text{PPO, applied}} = \begin{cases} u_k^{\text{PPO}}, & \text{with probability } 1 - p_{\text{drop}}, \\ 0, & \text{if packet dropped.} \end{cases} \quad (58)$$

E. DEEP Q-NETWORK (DQN)

The continuous action range is discretized into a finite set \mathcal{A} . The Q-network approximates

$$Q(s_k, a; \theta_q). \quad (59)$$

Action selection is ϵ -greedy:

$$a_k^{\text{DQN}} = \begin{cases} \arg \max_{a \in \mathcal{A}} Q(\tilde{s}_k, a), & \text{with prob. } 1 - \epsilon, \\ \text{random element of } \mathcal{A}, & \text{with prob. } \epsilon. \end{cases} \quad (60)$$

The TD update is

$$Q_{\theta_q}(s_k, a_k) \leftarrow r_k + \gamma \max_{a'} Q_{\theta_q}(s_{k+1}, a'). \quad (61)$$

F. ACTUATION MODEL AND CYBER NOISE

The physically applied action includes control noise representing edge uncertainty:

$$u_k^{\text{applied}} = u_k + e_k, \quad e_k \sim \mathcal{N}(0, \sigma_{\text{edge}}^2), \quad (62)$$

The grid dynamics follow the differential-algebraic model

$$\dot{x} = f(x, y, u_k^{\text{applied}}), \quad (63)$$

$$0 = g(x, y), \quad (64)$$

which determines the next physical state s_{k+1} .

G. HYBRID CONTROLLER COORDINATION

At each bus and time step, the supervisory layer evaluates the instantaneous cost of each controller:

$$c_k^{\text{ADP}}, \quad c_k^{\text{PPO}}, \quad c_k^{\text{DQN}}, \quad (65)$$

computed using (48). The hybrid supervisor selects

$$u_k^* = \arg \min_{u \in \{u^{\text{ADP}}, u^{\text{PPO}}, u^{\text{DQN}}\}} c_k(u). \quad (66)$$

H. ALGORITHM

The proposed algorithm of AI Enhanced Hybrid Cyber-Physical Adaptive Control incorporates real time sensing, state estimation and adaptive decision making to optimise the performance of smart grid. It makes use of neural network based policies and Approximate Dynamic Programming (ADP) to iteratively generate and evaluate the control actions. At each time step, measurements from PMUs, SCADA and IoT sensors are processed to estimate the system state and candidate control signals are then generated and evaluated with respect to a value function. Learning is done through temporal-difference updates and reinforcement feedback so that the policy performance is continuously improved.

Data:

System graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$;

Dynamic states $x(t)$;

Measurements $y(t)$ from SCADA/PMUs;

Control objectives \mathcal{J} (e.g., cost minimization, stability);

Neural Network policy model $\pi_\theta(x)$;

Value function approximator $\hat{V}_\phi(x)$

Result: Intelligent control signals $u^*(t)$ to stabilize and optimize grid performance

Initialization:

Define physical dynamics $\dot{x}(t) = f(x, u)$;

Define cyber measurements $y = h(x)$;

Initialize neural network weights θ, ϕ ;

while grid is operational do

Sensing: Acquire $y(t)$ from PMUs, SCADA, and IoT sensors;

State Estimation: Estimate $\hat{x}(t)$ using Kalman filter or AI-based state predictor;

AI-Based Control Policy Update:

Use neural network policy $\pi_\theta(x)$ to generate candidate control actions;

ADP-Based Evaluation:

Compute optimal action using:

$$u^*(t) = \arg \min_u [r(x, u) + \gamma \hat{V}_\phi(x_{t+1})];$$

Learning:

Update $\hat{V}_\phi(x)$ using temporal difference learning;

Update policy $\pi_\theta(x)$ using reinforcement feedback (e.g., policy gradient, actor-critic);

Execution:

Apply $u^*(t)$ to grid actuators (generators, FACTS, inverters);

Resilience Evaluation:

Simulate perturbations or faults and adapt policy accordingly;

end

return Stable and intelligent dispatch trajectory across cyber-physical system

Algorithm 1: AI-Enhanced Hybrid Cyber-Physical Adaptive Control for Smart Grid

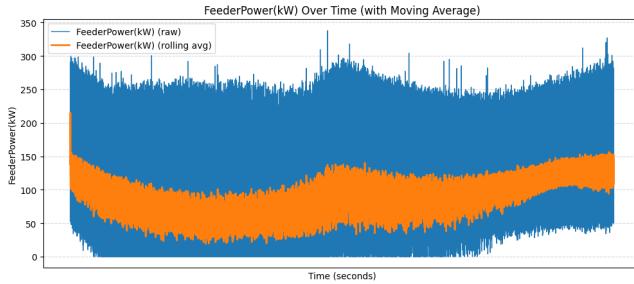


FIGURE 7: Voltage evolution at Bus 5 with local control actions.

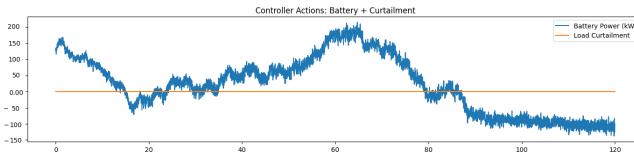


FIGURE 8: Per-unit Load, PV, and Wind power profile at Bus 5.

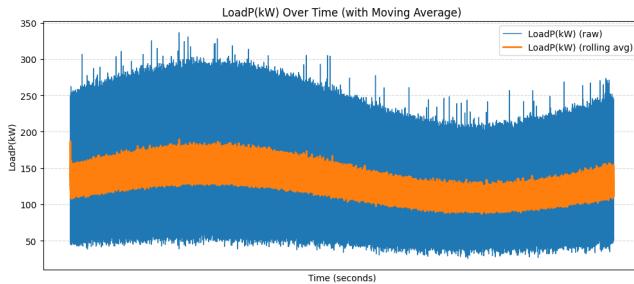


FIGURE 9: Temporal variation of load showing daily fluctuations and peak/off-peak demand patterns in the smart grid system.

bance events, confirming that the hybrid control framework minimizes disruptive corrective actions. Fig. 10 depicts the evolution of the average resilience index. The index remains predominantly in the high-performance band (0.95–1.0), despite occasional dips to 0.70–0.75 caused by rapid load transitions, cloud-induced PV drops, or FDI attacks. Fast recovery after each disturbance demonstrates the system's strong self-healing and adaptive capability under the hybrid control architecture. The updated control cost trajectories are shown in Fig. 11. The Total Control Cost fluctuates between 18 and 24, driven primarily by the system-wide corrective actions required during renewable and load disturbances. The PPO cost remains centered around 9, exhibiting stable and smooth behavior indicative of reliable cloud-level policy updates. The ADP and DQN edge controllers show lower costs, typically between 5 and 6, reflecting their fast response and low computational burden. Their trajectories track one another closely, confirming consistent learning behavior across edge devices. Occasional coordinated spikes

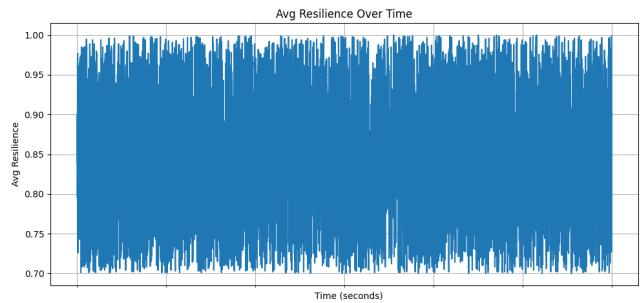


FIGURE 10: Average resilience index over the simulation horizon.

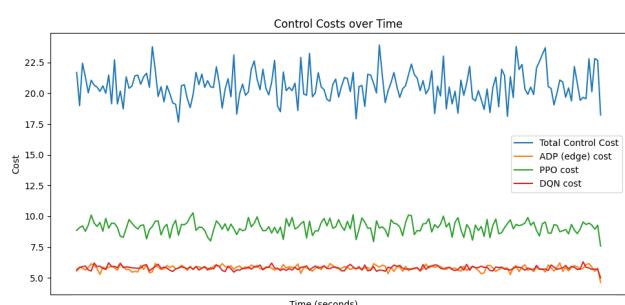


FIGURE 11: Control cost evolution for Total, ADP, PPO, and DQN controllers.

align with periods of heavy disturbance, but rapid recovery follows each event. Overall, the hybrid reinforcement learning architecture maintains high operational resilience despite cyber-physical disturbances. The cloud-level PPO agent ensures stable policy learning with minimal variability, while the ADP and DQN edge agents provide robust and fast real-time reactions. The strong coordination between cloud and edge layers results in stable voltages, controlled feeder flows, and low corrective action requirements. The system consistently demonstrates self-healing capability, confirming the suitability of the proposed framework for real-world smart grid applications.

VIII. CONCLUSION

The suggested hybrid cyber-physical model of smart grids has been already put into practice, and it proves the success of the 3-layer approach to the smart grid comprising physical, cyber, and control layers. The system has been found to be adaptive, scaled, and even sustainable by incorporating both Adaptive Dynamic Programming (ADP) and AI-based optimization approaches through cloud-independent as well as cloud-assisted contingencies. The use of simulation on a standard IEEE33 bus system has confirmed the fact that the framework is capable of providing grid stability, power dispatch optimization and responding effectively to dynamic operating conditions. By and large, the methodology proves to be an effective and viable solution to the smart grid

control and energy management of our current era with the successful implementation. Future development will be to scale the framework to incorporate real time hardware-in-the-loop testing and integration with emerging distributed energy resources to be more resilient and performant.

ACKNOWLEDGMENT

The authors would like to sincerely acknowledge the computing lab staff of NFC Institute of Engineering and Technology (IET), Multan, for their technical assistance and support during this research work.

REFERENCES

- [1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 944–980, 2012.
- [2] M. Amin and B. Wollenberg, "Toward a smart grid: power delivery for the 21st century," *IEEE Power and Energy Magazine*, vol. 3, no. 5, pp. 34–41, 2005.
- [3] W. Zhang, R. Yu, Y. He, and L. Huang, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 86–105, 2012.
- [4] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 24–45, 2015.
- [5] J. Gao, Y. Xiao, and J. Liu, "Cyber-physical security of a smart grid infrastructure," *Security and Communication Networks*, vol. 5, no. 8, pp. 825–834, 2012.
- [6] C. Liu, N. Yu, and J. Wang, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Applied Energy*, vol. 253, p. 113596, 2019.
- [7] L. Chen, J. Zhao, W. Lin, and P. Li, "Deep reinforcement learning for smart grid: A review of applications and challenges," *Renewable and Sustainable Energy Reviews*, vol. 149, p. 111498, 2021.
- [8] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game theory for demand side management in smart grid," *IEEE Communications Magazine*, vol. 50, no. 5, pp. 86–93, 2012.
- [9] S. Sharma and S. C. Srivastava, "Game-theoretic applications in electricity markets: A review," *Electric Power Systems Research*, vol. 158, pp. 148–159, 2018.
- [10] H. Farhangi, "The path of the smart grid," *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18–28, 2010.
- [11] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: Communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, 2011.
- [12] A.-H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 320–331, 2010.
- [13] J. A. Momoh, "Smart grid design for efficient and flexible power networks operation and control," *Power Systems*, vol. 23, no. 3, pp. 633–638, 2012.
- [14] S. Chakraborty, A. Hoke, M. Reno, B. Lundstrom, and M. Baggu, "Power system flexibility: A review of initiatives and technologies for wind integration," National Renewable Energy Laboratory (NREL), Golden, CO, 2016.
- [15] R. Yang, X. Wang, and J. Chen, "Distributed control for economic dispatch of smart microgrid with flexible loads," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4184–4193, 2019.
- [16] X. Liu, X. Hu, and X. Li, "Cyber security and privacy issues in smart grids," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 981–997, 2009.
- [17] Q. Huang, J. Yan, and W. Li, "A review of cyber-physical attacks and countermeasures in the smart grid," *IEEE Access*, vol. 5, pp. 24788–24806, 2018.
- [18] IEEE Standards Association, "IEEE Guide for Smart Grid Interoperability of Energy Technology and Information Technology Operation with the Electric Power System (EPS)," IEEE Standard 2030-2011. [Online]. Available: <https://standards.ieee.org/standard/2030-2011.html>
- [19] IEEE Standards Association, "IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces," IEEE Standard 1547-2018. [Online]. Available: <https://standards.ieee.org/standard/1547-2018.html>
- [20] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 998–1010, 2013.
- [21] A. Khanna and R. Kaur, "Energy management in smart cities based on internet of things: Peer-to-peer electricity trading using blockchains," *Energy Reports*, vol. 2, pp. 137–143, 2016.
- [22] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, 2011.
- [23] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, pp. 1–33, 2011.
- [24] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.
- [25] D. B. Rawat and C. Yu, "Cybersecurity in smart grid: Survey and challenges," *Computer Networks*, vol. 57, no. 5, pp. 1344–1371, 2013.
- [26] T. Alam, M. B. Mollah, and S. Islam, "Fog computing and its role in the internet of things," *International Journal of Computer Applications*, vol. 180, no. 7, pp. 1–7, 2017.
- [27] F. Tao, W. Zhao, and Q. Liu, "Edge intelligence: The confluence of edge computing and artificial intelligence," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4579–4590, 2019.
- [28] V. Kulkarni, H. Gill, and H. Sivakumar, "Edge computing in smart grids: Survey and challenges," *IEEE Access*, vol. 7, pp. 164467–164483, 2019.
- [29] S. Mohagheghi, J. Stoupis, C. Edrington, and M. Simões, "Distributed intelligent agents for autonomous microgrid management," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 57–65, 2010.
- [30] NIST, "Smart Grid Interoperability Panel: Framework and roadmap," Special Publication 1108r3, 2014.
- [31] M. Amin and B. Wollenberg, "Smart grid: Overview, issues and opportunities," *Journal of Advanced Research*, vol. 1, no. 1, pp. 1–9, 2011.
- [32] Y. Yang, X. Wang, and A. V. Vasilakos, "A survey on smart grid communication infrastructures: Motivations, requirements and challenges," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 5–20, 2013.
- [33] H. Sun, J. Li, and B. Wang, "Review of smart grid technologies: Potential applications and challenges," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1486–1511, 2018.
- [34] T. Salman and R. Jain, "Review of blockchain-based energy trading systems: Current status and research challenges," *Electronics*, vol. 9, no. 4, p. 530, 2020.
- [35] J. Wang, C. Shen, and F. Liu, "Blockchain-based smart contracts for decentralized energy management of distributed energy resources in smart grid," *IEEE Access*, vol. 7, pp. 118906–118917, 2019.
- [36] N. Rezaei and M. Dabbagh, "Distributed ledger technologies for smart energy markets: A review of blockchain-based peer-to-peer energy trading platforms," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109547, 2020.



MUHAMMAD SIDDIQUE has received the B.Sc. degree in Electrical Engineering from BZU, Multan, Pakistan, in 2006, and the M.Sc. and PhD degrees in Electrical Engineering from PIEAS, Pakistan, in 2007 and 2018 respectively. In 2023, he had completed his postdoc from Yale University USA. He is currently working as HOD and active researcher in Artificial Intelligence department focusing on smart grids, Cyber-physical systems and AI based adaptive control.

He has been working as a Lecturer and Assistant Professor at NFC IET Multan, where he is involved in teaching, research and laboratory development of power systems and renewable energy integration. His recent work has focused on smart grid security, machine learning applications for energy systems and distributed energy resources. He has published a large number of papers in the journals and conferences by the Institute of Electrical and Electronics Engineering and Elsevier on these topics. He has been a participant in many conferences and has been acknowledged for his contributions to the field of smart grid research and has worked with international researchers in AI in power systems. Dr. Siddique is has supervised many undergraduate and post graduate students for their research work.



SOHAIB ZAFAR Received his B.Sc Electrical Engineering Degree from University of Engineering and Technology, Lahore in 2014, and MS in Electrical Engineering from Lahore University of Management Sciences. His research interests include renewable energy systems, modeling and optimisation, smart grid and power systems and control. He has previously worked as Research Associate at Energy Informatics Group at Lahore University of Management Sciences as part of National Center for Big Data and Cloud Computing, Pakistan. He is an aspiring academic in the field of energy and power system.ty, College Station. He is the author of three books, more than 150 articles.

• • •