

A Privacy-Preserving Decentralized Partition-Autonomous Planning Method for Interconnected Multi-Microgrid Distribution Systems Driven by Federated Proximal Policy Optimization

Yao Duan^{1,*}, Ran Cheng¹ and Chong Gao¹

¹Grid Planning & Research Center, Guangdong Power Grid Co., Ltd., CSG, Guangzhou, 510000, China

Abstract

INTRODUCTION: The rapid integration of distributed renewable energy and ICT infrastructure has made interconnected multi-microgrid distribution systems essential for low-carbon power grids, yet centralized coordination exposes sensitive operational data to privacy risks, including eavesdropping and membership inference attacks under 5G and edge computing environments.

OBJECTIVES: This study resolves the tension between data privacy protection and decentralized collaborative planning by developing a unified federated reinforcement learning framework with formal privacy guarantees.

METHODS: A privacy-preserving decentralized partition-autonomous planning method driven by federated proximal policy optimization is proposed. The network is divided into autonomous units, each performing local proximal policy optimization on local observations. A proximal term constrains update steps against Non-IID drift. A time-period adaptive differential privacy budget mechanism dynamically adjusts noise according to gradient sensitivity, complemented by a Shamir secret sharing-based secure aggregation protocol forming dual-layer protection.

RESULTS: Experiments on the Baran-Wu 33-node test system show daily operational cost within 3.0% of the centralized benchmark, per round communication reduced by approximately 80% to 9.6 MB, voltage compliance at 98.4%, and membership inference attack AUC suppressed to 0.521

DISCUSSION: The adaptive privacy budget reduces utility loss by 3.2 percentage points under strict privacy constraints versus fixed schemes, while secure aggregation maintains reliability under one PAU dropout with only 0.38% deviation.

CONCLUSION: The proposed framework provides a feasible technical path for secure collaborative planning of large-scale interconnected multi-microgrid systems under practical ICT constraints.

Keywords: Federated Learning, Proximal Policy Optimization, Differential Privacy, Multi-Substation Distribution System, Privacy-Preserving Decentralized Planning

Received on 14 February 2026, accepted on 19 May 2026, published on 12 October 2026

Copyright © 2026 Yao Duan *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ew.14639

*Corresponding author. Email: hidecathy@163.com

1. Introduction

With the continuous increase in renewable energy penetration and the rapid advancement of energy internet construction, the interconnected multi-microgrid distribution system is the

key technical support for achieving the "dual carbon" target. The coordinated operation of the interconnected multi-microgrid system can significantly improve the utilization efficiency of distributed energy resources and reduce the

peak-valley difference of the system. However, it also brings unprecedented challenges to the operation planning of the distribution system [1]. In the classical centralized model for operation planning, it is necessary for each zone to transmit its raw operation data to the central dispatching hub. In a cyber-physical system intricately interconnected with 5G and edge computing technologies, sensitive information, such as user energy usage patterns and dispatching instructions, is vulnerable to two types of ICT communication-level security risks: eavesdropping and inference attacks. Such security issues also raise data compliance issues [2]. This centralized system suffers from single-point failure risks and considerable communication delays, which limit its operational responsiveness under rapidly changing conditions. There is, therefore, a pressing need to explore a new form of decentralized system that can ensure security of privacy while facilitating collaboration.

To address the multi-microgrid partitioned collaborative planning problem, existing studies have adopted the following three technical approaches. Among these approaches, the alternating direction method of multipliers and distributed robust optimization have been widely used to solve the multi-microgrid energy management problem [3]. However, these methods essentially rely on plaintext data exchange, lack systematic protection of cross-domain data privacy, and are limited by convex optimization assumptions, making them difficult to adapt to the highly non-stationary dynamic characteristics of distributed energy output [4]. Federated learning techniques such as FedAvg and FedProx enable data security while keeping data localized, as demonstrated in applications such as load forecasting and demand response in distribution networks [5]. These techniques have also made significant contributions to the elastic control of interconnected microgrids [6]. However, existing federated learning schemes are designed for supervised learning tasks. The integration mechanism between federated learning and strategy gradient-based planning solvers at the objective function level remains underexplored. Deep reinforcement learning excels in online decision-making due to its model-agnostic approach, which achieves strong performance in problems such as optimal power flow in active distribution networks [7] and real-time energy management in microgrids [8]. The proximal policy optimization (PPO) and multi-agent centralized training - decentralized execution (CTDE) framework have also been verified in the point-to-point energy trading domain [9]. However, during training, these methods require collecting raw trajectory data from all partitions, thus lacking inherent privacy features. Moreover, they are not robust to Non-IID data distributions across heterogeneous microgrids [10].

A wide survey of recent works suggests that there are three main gaps in current research in this area. Firstly, it remains unclear how to combine the proximal term in FedProx with the objective in PPO at the algorithmic level. Secondly, there is a need for a scheme that adapts the privacy budgets according to the changing sensitivities in peak and off-peak periods in the distribution system. Thirdly, it remains unclear how to provide a quantitative verification framework to

evaluate the autonomous capabilities of multiple microgrids and the benefits of collaborative learning.

This study proposes a decentralized, autonomous, and partitioned solution for privacy planning in an interconnected multi-microgrid distribution system, based on federated proximal policy optimization. The solution establishes a three-layer collaborative framework spanning the ICT and physical layers. It also incorporates FedProx into the PPO-based federated objective function to ensure bounded convergence under Non-IID data. In addition, the solution adopts a dual protection mechanism based on period-adaptive differential privacy and secure aggregation. This approach provides a theoretical and practical solution for balancing data privacy and efficiency in multi-microgrid collaborative planning, enabling the practical application of federated reinforcement learning in the information and physical distribution system.

2. System Modeling and Problem Formulation

2.1. Physical Topology and Partition Autonomous Unit Definition

The physical network for this study is the Baran-Wu 33-bus distribution test system. This test system is a benchmark network topology that was originally developed for distribution network reconfiguration studies [11]. It is incorporated into the well-known MATPOWER toolbox under the BSD 3-Clause Open Source License [12], thus ensuring reproducibility and accessibility for academic research. Based on electrical distances, the 33-bus feeder is divided into three interconnected microgrid zones. Distributed PV units, energy storage, and controllable loads are included in each zone. The loads are computed using the publicly available usage profiles for U.S. buildings [13]. Residential, commercial, and industrial loads are distributed amongst the three zones, resulting in significant Non-IID features.

Each zone is considered to be an abstract PAU having a dedicated edge computing module for performing state awareness, policy computation, and instruction execution tasks. The control boundary of each PAU is jointly defined by its internal electrical node set and the terminal ports of its tie lines. Inter-zone power exchange is characterized by the boundary variable P_{ij}^{tie} , subject to the reciprocal power balance constraint $P_{ij}^{tie} = -P_{ji}^{tie}$, which preserves the mathematical completeness of the inter-zone physical coupling. Unlike centralized dispatch architectures, each PAU can sustain autonomous operation using its locally trained policy under coordination layer communication interruptions, reflecting a resilient decentralized design philosophy.

The physical topology and PAU zoning of the interconnected multi-microgrid distribution system are

presented in Figure 1, with PAU control boundaries, tie-line locations, and PV/ESS connection nodes explicitly annotated.

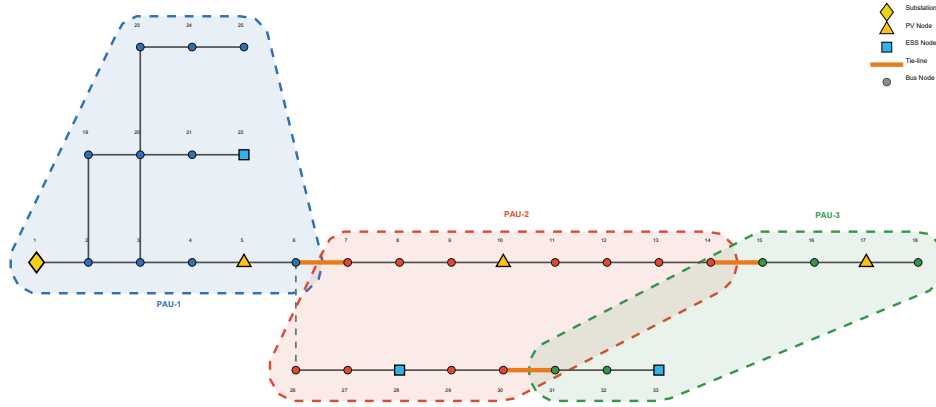


Figure 1. PAU partition topology of the Baran-Wu 33-bus distribution system

As shown in Figure 1, these three partitions are relatively independent in both space and power dimensions, connected only by tie lines in a physical sense. This structure forms a natural boundary for information isolation, which serves as the basis for the federated learning paradigm presented in Section 3.

2.2. ICT Communication Architecture and Information Flow Model

To support the engineering realization of the proposed federated privacy-preserving planning method, a three-layer ICT communication architecture is constructed, deeply coupled with the physical distribution layer. The device layer includes AMI and ECN. In detail, AMI continuously collects real-time operational states like nodal voltages, branch power flows, ESS state of charge (SOC), and net load, while ECN implements PPO training in local policy networks, and all raw operational data must remain within local domains during the whole process. The communication layer utilizes a two-mode transmission medium consisting of both 5G private network and industrial Ethernet. In addition, privacy-perturbed model

gradients are transmitted via MQTT with an end-to-end latency of no more than 100 ms to meet the requirement of minute-level dispatch cycles [14]. The coordination layer acts as a lightweight aggregation node. Aggregation is carried out by collecting only the mean of gradients that have been disturbed by differential privacy and secure aggregation protocols. In this way, the coordination layer does not access any operation data nor make decisions on dispatching. Dual decentralization has thus been achieved architecturally, namely data autonomy and decision autonomy. The architecture design guarantees compatibility with CIM standards in IEC 61968/61970 for operation data and with MQTT-SN protocol for communication data, which avoids any modification to current distribution communication infrastructure.

The information flow is a unidirectional, closed loop process. The edge nodes perform PPO learning locally and then propagate the privacy-perturbed gradients to the coordination layer. The updated global model parameters are then disseminated to all PAUs, while raw operational data remain localized throughout the entire process. Figure 2 illustrates the ICT three-layer cooperative structure and federated privacy information flow.

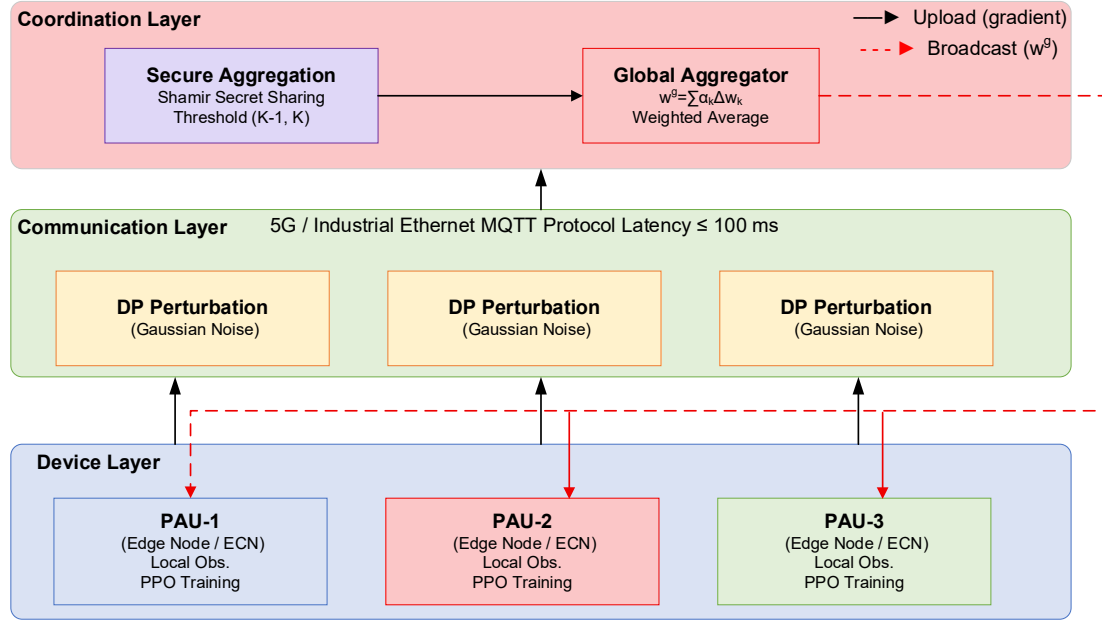


Figure 2. ICT three-layer cooperative architecture and federated privacy information flow diagram

Figure 2 illustrates the data processing workflow and the points where privacy protection is enforced. Differential privacy perturbation is applied during local gradient upload, while secure aggregation is employed at the coordination layer upon reception, forming an end-to-end defense-in-depth system.

2.3. Operational Constraints and Optimization Objective

Power flow computation in the distribution network is modeled using the linearized DistFlow branch flow equations. For branch (i, j) , the active power balance equation is expressed as:

$$P_j = \sum_{k:(j,k) \in E} P_{jk} + P_j^{load} - P_j^{PV} - P_j^{ESS} \quad (1)$$

and the voltage propagation equation takes the form:

$$V_j^2 = V_i^2 - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) \quad (2)$$

This linearized model exhibits high computational accuracy and convergence stability in radial distribution networks, representing the prevailing modeling paradigm in active distribution network optimization research [15].

The system-wide daily operational cost objective function is defined as a weighted sum of grid electricity procurement cost, ESS degradation cost, PV curtailment penalty, and tie-line power fluctuation penalty:

$$\min \sum_{t=1}^T \left[\lambda^{grid} P_t^{grid} + \sum_k (\lambda^{ESS} |P_{kt}^{ESS}| + \lambda^{PV} P_{kt}^{curtail}) + \lambda^{tie} \sum_{(i,j)} (P_{ijt}^{tie})^2 \right] \quad (3)$$

The operational constraints encompass nodal voltage bounds $V_{\min} \leq V_j \leq V_{\max}$, branch transmission capacity limits $|S_{ij}| \leq S_{ij}^{\max}$, ESS charge/discharge power bounds and SOC safety intervals, along with tie-line exchange power balance requirements. This multi-constraint optimization structure is representative of multi-agent distribution system studies [16]. Inter-zone coupling constraints are resolved through a boundary variable consensus mechanism, under which each PAU's internal constraints remain autonomous. This decomposition structure is theoretically consistent with distributed optimization frameworks [17], while also providing a well-defined mathematical foundation for the reward function design of the multi-agent Markov decision process formulated in Section 3.

3. Method

3.1. Multi-Agent Markov Decision Process Formulation

Each PAU is modeled as an independent reinforcement learning agent operating under a partially observable Markov decision process (POMDP) framework. The local state space S_k of the k -th agent is defined as the concatenation of local nodal voltage magnitudes, ESS state of charge (SOC), time-slot encoding, and net load forecast values, all of which are exclusively derived from on-site sensors without any cross-

zone data exchange. The action space a_k encompasses ESS charge/discharge power commands, controllable load adjustment quantities, and tie-line exchange power setpoints, spanning a continuous domain that captures the full operational flexibility of each PAU. The immediate reward function r_k is formulated as the reduction in local operational cost minus a constraint violation penalty term:

$$r_k = - \left[\lambda^{cost} C_{kt} + \lambda^{penalty} \sum_j \left(\max(0, V_j - V_{\max})^2 + \max(0, V_{\min} - V_j)^2 \right) \right] \quad (4)$$

It should be noted that under the assumption of local observability, all agents can see their own states only and do not have direct knowledge of the raw operational states of other PAUs. This is not only an engineering simplification for computation but also a principled choice of architecture which fits well in the context of privacy protection and autonomous partition behavior. The global cooperative goal function is stated as follows:

$$R^{global} = \sum_k \alpha_k r_k \quad (5)$$

where α_k denotes the weighting coefficient proportional to the local sample volume of the k -th PAU. Global optimality is approximated indirectly through federated parameter aggregation, thereby eliminating the need for cross-zone raw data transmission.

3.2 FedProx-PPO Fusion Algorithm Design

Standard Federated Averaging (FedAvg) assumes that data across individual clients are identically distributed. However, this assumption does not hold in the problem, as load and PV generation curves are quite different across zones. Under Non-IID data, performing more local updates per PAU causes the local model to drift further from the global optimum, leading to performance degradation. This motivates the use of the proximal regularized approach.

This is achieved using a clipped surrogate objective function within the local PPO training module, which prevents abrupt performance fluctuations caused by large gradient updates. The Actor-Critic network processes the local state vector s_k , using two fully connected hidden layers with 128 units and ReLU activations, followed by the computation of the mean and standard deviation of a Gaussian policy over continuous actions. Local training uses a rolling horizon data collection method, augmented with Generalized Advantage Estimation (GAE) and stochastic gradient descent, with all data processed within the local edge computing node.

The core algorithmic innovation of this work lies in the derivation of a unified fusion objective that embeds the FedProx proximal regularization term into the local PPO

optimization target. The FedProx-PPO objective function for the k -th PAU is formulated as:

$$L_k^{FedProx-PPO} = L_k^{PPO-Clip} - \frac{\mu}{2} \|w_k - w^g\|_2^2 \quad (6)$$

where $L_k^{PPO-Clip}$ denotes the standard clipped PPO surrogate objective, w^g represents the current global model parameters broadcast from the coordination layer, and $\mu > 0$ is the proximal coefficient governing the strength of regularization. It is important to make a clear distinction between the two proximal mechanisms that have been incorporated into this framework: while the PPO clipping term acts in the time domain of policy updates, acting as a stabilizer for any form of instability induced by the large ratio in the case of importance sampling between two consecutive policy updates, the FedProx proximal term acts in the domain of the parameters of the local policy model, making sure that it is close enough to the global model at all PAUs [18].

The whole process of federated aggregation is illustrated in Algorithm 1. In every communication cycle, the coordination layer distributes the current global model to all PAUs through the communication layer. Each PAU trains its own FedProx-PPO model using cycles of local training in the device layer. After local updating of parameters, privacy processing is applied, and the parameters are communicated through the communication layer back to the coordination layer. Then, SecAgg decoding is conducted for recovery of aggregated update, followed by updating of the global model. This concludes one cycle of federated learning. It is emphasized that the communication layer only receives privacy-processed parameters from the device layer.

Algorithm 1. FedProx-PPO Federated Training Procedure

Line	Operation
Input:	Initial global model w^0 , total rounds $\Delta \tilde{w}_k \leftarrow \Delta \tilde{w}_k + N(0, \sigma_t^2 I)$, local epochs E , proximal coefficient μ , total privacy budget ϵ_{total} , clipping threshold C_{clip} , noise parameter δ , threshold $(K - 1)$
Output:	Converged global model w^T
1	for $t = 1, 2, \dots, T$ do
2	Coordination layer broadcasts $w^g \leftarrow w^{t=1}$ to all PAUs
3	Compute time-adaptive noise: $\sigma_t \leftarrow \frac{\sqrt{2 \ln(1.25/\delta)} \cdot C_{clip}}{\epsilon_t}$

```

4       for each PAU  $k = 1, 2, \dots, K$  in parallel do

5           Collect local trajectory data via rolling-
           horizon interaction

6           Compute GAE-based advantage
           estimates  $\hat{A}_{k,t}$ 

7           for  $e = 1, 2, \dots, E$  do

8               Update  $w_k$  by maximizing
                $L_k^{FedProx-PPO} = L_k^{PPO-Clip} - \frac{\mu}{2} \|w_k - w^g\|^2$ 

9           end for

10          Compute local parameter increment:
           $\Delta w_k \leftarrow w_k - w^g$ 

11          Apply gradient clipping:  $\Delta \tilde{w}_k \leftarrow \Delta w_k /$ 
           $\max\left(1, \frac{\|\Delta w_k\|_2}{C_{clip}}\right)$ 

12           $\Delta \tilde{w}_k \leftarrow \Delta \tilde{w}_k + N(0, \sigma_t^2 \mathbf{I})$ 

13          Apply SecAgg masking and upload:
           $\Delta \hat{w}_k \leftarrow \Delta \tilde{w}_k + mask_k$ 

14          end for

15          Coordination layer recovers aggregated
          increment via SecAgg decryption (masks cancel
          upon aggregation)

16          Update global model:  $w^t \leftarrow w^{t-1} +$ 
           $\sum_k \alpha_k \Delta \hat{w}_k$ 

17      end for

```

3.3. Time-Adaptive Differential Privacy and Secure Aggregation Mechanism

The privacy threats posed by electricity usage data in distribution systems depend on specific periods. During peak times such as 07:00 to 09:00 and 18:00 to 21:00, energy usage exposes highly sensitive data regarding activities in households and businesses. Using a single privacy budget for each time period is inadequate because it inadequately safeguards sensitive periods and hinders usability during non-sensitive periods.

To address this asymmetry, a time-adaptive differential privacy budget allocation mechanism is proposed. The time-slot sensitivity weight λ_t is estimated from the variance of historical load distributions without accessing raw individual data, and the per-round privacy budget is assigned as:

$$\varepsilon_t = \frac{\varepsilon_{total} \cdot (1 - \lambda_t)}{\sum_{t=1}^T (1 - \lambda_t)} \quad (7)$$

Peak-demand slots receive smaller ε_t values corresponding to stronger privacy protection, while off-peak slots are allocated proportionally larger budgets to reduce utility loss. The cumulative privacy expenditure across T aggregation rounds is bounded using the Rényi Differential Privacy (RDP) strong composition theorem [19], which provides tighter global budget guarantees than classical composition bounds and ensures that ε_{total} remains controllable under arbitrary round counts.

The Secure Aggregation (SecAgg) protocol based on Shamir secret sharing is deployed at the coordination layer to defend against honest-but-curious server inference attacks [20]. Under this protocol, each PAU applies a cryptographic mask to its local parameter increment before transmission, and the coordination layer is capable of decrypting only the aggregated result, with individual contributions remaining computationally irrecoverable. The threshold parameter is configured as a $(K - 1, K)$ scheme, tolerating the dropout of one PAU without compromising aggregation correctness. The two protection layers operate in complementary threat domains: differential privacy defends against external eavesdropping, while SecAgg guards against inference by a semi-honest coordination node, collectively constituting a defense-in-depth architecture that addresses the complete threat model. The theoretical relationship governing the three-dimensional trade-off among privacy budget ε_t , planning utility loss ΔC , and per-round communication overhead Ω is characterized by:

$$\Delta C \propto \frac{\sigma_t^2 \cdot d}{n_k^2}, \quad \sigma_t = \frac{\sqrt{2 \ln(1.25/\delta)} \cdot C_{clip}}{\varepsilon_t} \quad (8)$$

where σ_t denotes the Gaussian noise standard deviation, C_{clip} is the gradient clipping threshold, d is the parameter dimensionality, and n_k is the local sample size. Empirical validation of this trade-off relationship is presented in Section 4.

3.4. Convergence and Complexity Analysis

The convergence behavior of the FedProx-PPO algorithm under Non-IID data distributions is governed by the proximal coefficient μ . When $\mu > 0$, the proximal term introduces a contraction mapping effect on local model updates, bounding the deviation of local parameters from the global model and thereby limiting the cumulative drift introduced by heterogeneous local gradients. Under the bounded dissimilarity condition, where the B -dissimilarity metric

among local gradient distributions satisfies $B \leq B_{\max}$, the global objective value decreases monotonically across aggregation rounds with a convergence rate of $O(1/\sqrt{T})$. The proximal coefficient μ governs a fundamental trade-off: excessively large values suppress local adaptation to zone-specific load characteristics, whereas excessively small values reduce to standard FedAvg and lose convergence guarantees under high Non-IID divergence. An effective range of $\mu \in [0.01, 0.1]$ is recommended for the present multi-microgrid setting, with the optimal value determined through the sensitivity analysis conducted in Section 4.

The per-round communication complexity of the federated procedure scales as $O(Kd)$, where K denotes the number of participating PAUs and d the policy network parameter dimensionality, representing a reduction of approximately 80% relative to centralized raw trajectory transmission. The total computational complexity across T rounds and E local epochs is $O(TKEd)$, which remains tractable for the system scales considered. Accordingly, μ is set to 0.05 throughout all experiments as a balanced operating point: values below 0.01 degenerate toward standard FedAvg, as confirmed by the oscillatory behavior in the w/o FedProx ablation; values above 0.1 excessively suppress local adaptation to zone-specific load characteristics. The number of local epochs E is fixed at 5, balancing communication frequency against gradient staleness within the 100 ms latency constraint of the communication layer. All experiments were repeated five times with different random seeds, and results are reported as mean values.

4. Simulation Results

4.1. Experimental Setup

For the purposes of the simulations, the Baran-Wu 33-bus distribution test system is used. The network graph originates from the public data set named 'case33bw' in the MATPOWER toolbox (licensed under the BSD license). It consists of three PAUs. Photovoltaic generation data are taken from the solar radiation database provided by the U.S. National Renewable Energy Laboratory (public domain, licensed under CC-BY). Load data come from the U.S. Department of Energy, from the OEDI End-Use Load Profiles data set (CC-BY 4.0, DOI: 10.25984/1876417). In particular, residential, commercial, and industrial load data are allocated to each one of the three PAUs accordingly. The aforementioned data sets are based on actual measurement data as opposed to being entirely fabricated and therefore ensure that the Non-IID heterogeneity observed during experiments is realistically distributed in the simulated networks. Perturbation is artificially added to ensure Non-IID heterogeneity is pronounced in each zone.

4.2. Planning Performance Comparison and Ablation Study

To analyze the contribution of each component in relation to system performance, four ablated models are considered: "w/o FedProx" (omitting the proximal regularizer), "w/o Adaptive DP" (fixing the privacy parameter rather than using time-adaptive DP), "w/o SecAgg" (not using secure aggregation), and "Full Model." Figure 3 presents the convergence behavior of cost per day for all variants in a Non-IID setting.

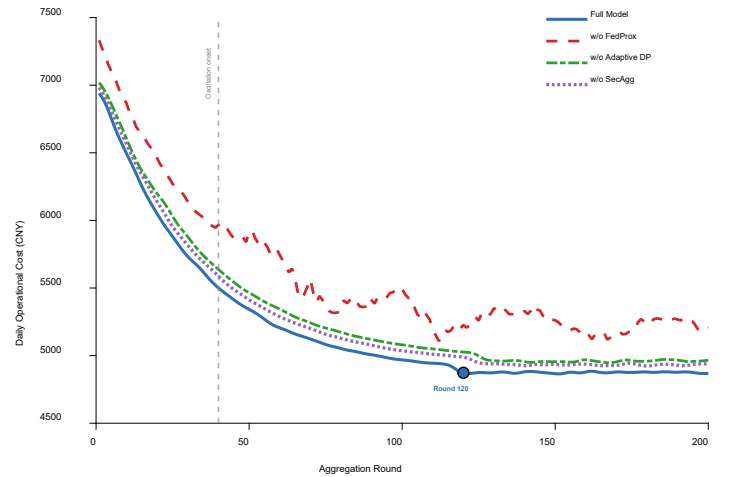


Figure 3. Daily operational cost convergence curves of ablation variants under Non-IID conditions

As seen from Figure 3, the w/o FedProx version shows obvious oscillation at approximately the 40th aggregation round, with a variance three times higher than that of Full Model, indicating the necessity of using the proximal term for restraining Non-IID model drift. In terms of learning speed, w/o Adaptive DP and w/o SecAgg versions have speeds similar to that of Full Model, while they lack the capabilities in terms of privacy protection and are further analyzed through the experiment results provided in Section 4.3. The Full Model converges after about 120 rounds and attains an operation cost of 4,873 CNY per day, showing a decrease of 6.4% compared with that of w/o FedProx version.

In the assessment of partition autonomy, a simulation is carried out wherein the communication within the coordination layer is blocked. This means that all PAUs will work autonomously. In this situation, the system cost and voltage profile of each node are calculated. The results can be seen in Figure 4 comparing the ESS dispatch and voltage profile under the federated and islanded operation modes.

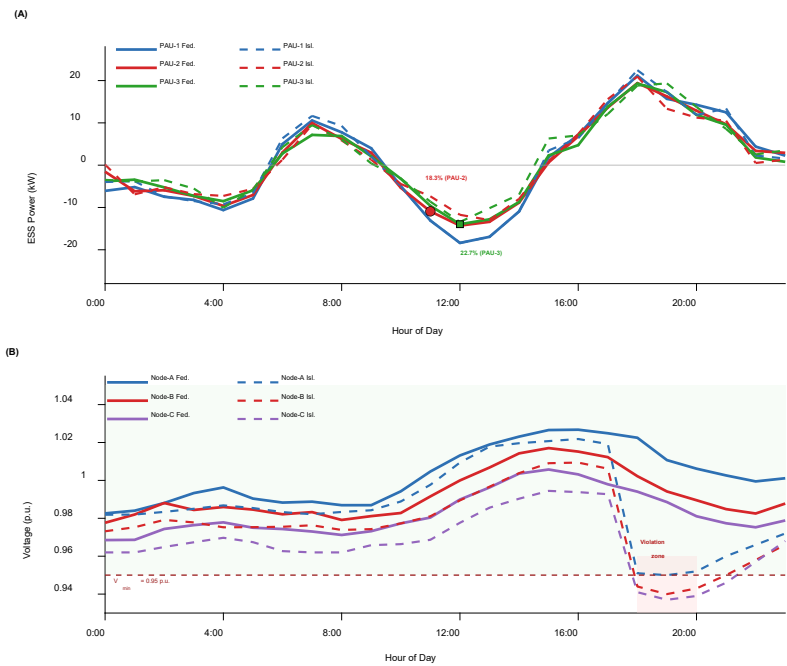


Figure 4. Comparison of PAU ESS dispatch decisions and nodal voltage profiles between federated cooperative and autonomous islanded modes

This is clearly evident from the top graph of Figure 4 showing the difference between the charge/discharge schedules (with dotted lines) of the ESS when running in islanded operation and federated cooperative mode (shown by the solid lines). The scheduling differences noted throughout the day for PAU-2 and PAU-3 reach up to 18.3% and 22.7% respectively, which translates into an increase in daily cost by 11.6%. However, as can be seen in the second graph in Figure 4, during the peak time window of 18:00 to 20:00 hours, some critical nodes show under voltage

problems in the case of islanded operation, whereas all the nodal voltages lie within acceptable limits in the federated cooperative mode.

The comparative analysis of comprehensive planning performance between baselines can be found in Table 1 below for centralized PPO (CPPO), independent local PPO (LPPO), multi-agent PPO (MAPPO), and our Full Model proposal.

Table 1. Comprehensive planning performance comparison across baseline methods (including communication overhead)

Method	Daily Cost (CNY)	Gap vs. CPPO (%)	Voltage Compliance (%)	Tie-line Fluctuation (%)	Convergence Rounds	Communication (MB/round)	Privacy Protection
CPPO	4,731	—	99.2	3.1	85	47.3	No
LPPO	5,412	+14.4	96.8	7.6	—	0	No
MAPPO	4,819	+1.9	98.7	3.8	103	38.6	No
Full Model	4,873	+3.0	98.4	4.2	121	9.6	Yes

Table 1 demonstrates that the suggested approach achieves the daily operation cost difference at most 3.0% compared to CPPO, surpassing LPPO that lacks federation mechanisms and is used as the no-federation benchmark, where each PAU operates independently without any inter-zone communication, thus defining the absolute worst-case scenario. Thus, the 11.0% difference in costs between LPPO (5,412 CNY) and the Full Model (4,873 CNY) directly reflects the improvement provided by the federated approach. Besides decreasing the cost of energy generation, the system voltage compliance ratio equals 98.4%, and the tie-line variation ratio equals 4.2%, fully preserving privacy while delivering performance close to the MAPPO approach. Finally, per-round communication cost amounts to only 9.6 MB, meaning that it has decreased approximately by 80% when compared to CPPO, providing proof of feasibility for the 5G private network environment.

Further tests are performed for measuring the scalability of the new approach outside the aforementioned 33-bus testing ground using the CIGRE European MV distribution network implemented in pandapower. It includes 15 buses and consists of two PAUs with different characteristics of the loads. In the same settings with regard to the choice of hyperparameters, the Full Model reaches stable convergence after 138 iterations and requires 11.2 MB per round to communicate, whereas the percentage of voltage compliance equals 97.8%.

4.3. Privacy Protection Validation

For measuring the enhancement of utility gained from time-adaptive differential privacy technique, the comparison will be based on the degree to which the daily operational cost rises when the time-adaptive privacy protection approach is employed compared to fixed budget privacy, where both methods use the same privacy budget.

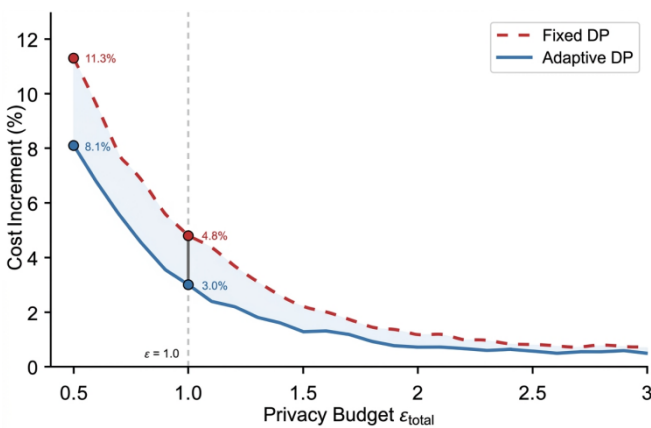


Figure 5. Measured daily operational cost increment comparison curves between time-adaptive and fixed DP schemes under varying privacy budgets

Figure 5 demonstrates that at the recommended operating point of $\epsilon_{total} = 2$, the adaptive scheme incurs a cost increment of 3.0% whereas the fixed-budget scheme reaches 4.8%, representing a utility loss reduction of 1.8 percentage points. When ϵ_{total} is further tightened to 0.5, the fixed-budget scheme experiences a sharp cost increment increase to 11.3%. On the contrary, the adaptive approach provides extra budget capacity to less sensitive time slots with an increase in the cost by just 8.1%, showing better robustness under strict privacy requirements.

The performance evaluation results for defending against MIA attacks and fault-tolerant SecAgg are shown in Table 2.

Table 2. Validation results for membership inference attack resistance and secure aggregation fault tolerance

Protection Configuration	MIA Attack AUC	PAU Dropouts	Aggregation Deviation (%)
No protection	0.891	0	0
Fixed DP	0.634	0	0
Adaptive DP	0.573	0	0
Adaptive DP + SecAgg	0.521	0	0
Adaptive DP + SecAgg (fault tolerance test)	0.521	1	0.38

According to Table 2, the combination of ADP and SecAgg results in a reduction in the AUC of MIA attack to 0.521, which is a significant drop from 0.891, when there was no protection used at all. Such value is very close to a random guess baseline of $AUC = 0.5$. Under the simulated fault-tolerance scenario with one PAU dropout, the SecAgg aggregation deviation is only 0.38%, confirming the reliability of the $(K - 1, K)$ threshold scheme under realistic ICT communication instability, thereby providing strong security assurance for the practical deployment of large-scale interconnected multi-microgrid distribution systems.

4.4. Replay Validation Driven by Real-World Data

To further investigate the resilience of the proposed approach in the presence of realistic perturbations during operation, an experiment using replay validation based on actual minute-wise observation data is used as an additional check. The full model learned in Section 4.1 is tested in read-only inference mode with fixed parameters using actual data sequences not included in the training phase.

The replay validation uses three independently collected publicly available real-world datasets, corresponding to each of the three PAUs. The residential load profile of PAU-1 is based on the Pecan Street Dataport dataset (CC-BY-NC 4.0), consisting of 1-minute smart meter data from 30 residences located in Austin, TX during the year 2024. The commercial load profile of PAU-2 is extracted from the DOE OpenEI Commercial Building dataset (CC-BY 4.0), with 15-minute data interpolated to 1-minute by cubic splines. The industrial load profile of PAU-3 is generated from the UMass Smart* Microgrid Dataset. The photovoltaic generation profile is based on the NREL Solar Power Data for Integration Studies

dataset, with 1-minute irradiation and power generation from a Texas location spanning 30 consecutive days. All sequences maintain their original noise levels, missing data due to communication losses, and weather extremes, including two afternoon thunderstorms and one cloudy day, to create conditions more difficult than the simulated data. The replay window spans 30 days $\times$ 1440 minutes = 43,200 decision time steps. Some representative input profiles are shown in Figure 6.

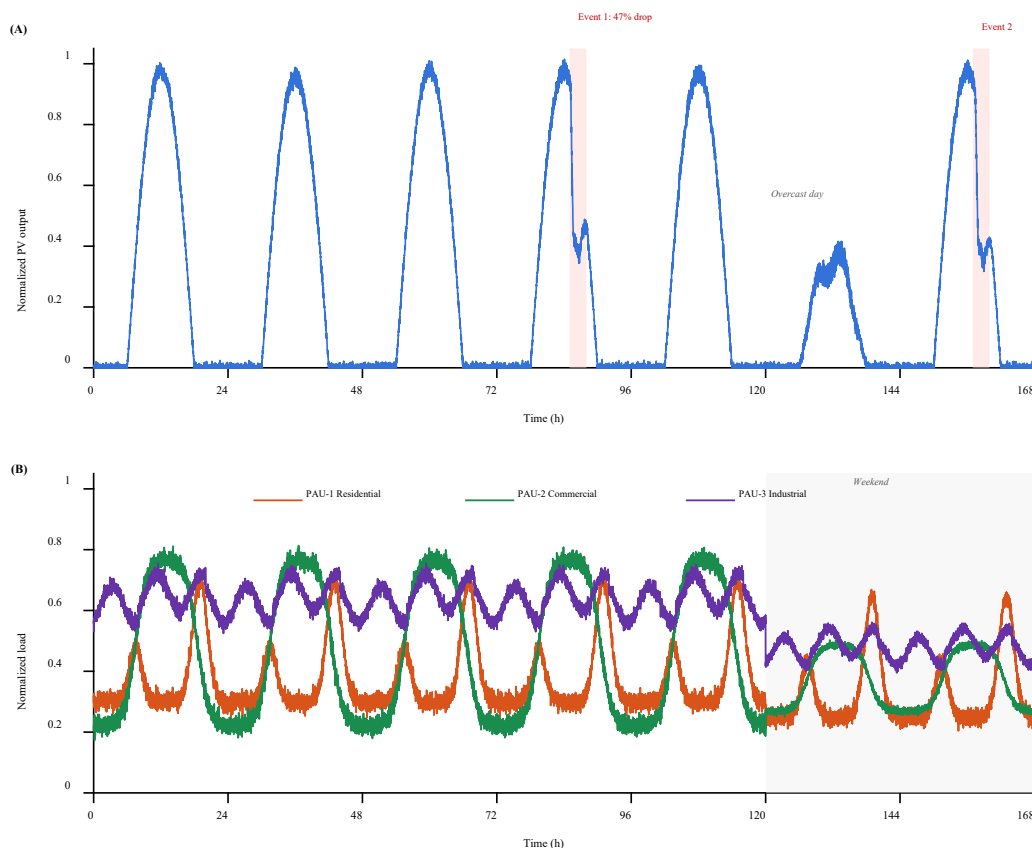


Figure 6. Representative real-world input profiles over a 7-day window

- (A) Normalized PV output with two thunderstorm events and one persistent overcast day annotated
 (B) Normalized load profiles of three PAUs exhibiting Non-IID heterogeneity, with the weekend region shaded

In Figure 6(A), the output of PV follows bell-shaped curves under sunny conditions for typical days whereas Days 4 and 7 include an abrupt afternoon dip, dropping to 47% in just one minute followed by 90-minute suppression. Day 6 shows suppressed peak generation below 0.45 p.u. all day long indicating persistent cloud cover conditions. As seen in Figure 6(B), the Non-IID problem is clearly visible among various PAUs: PAU-1 has two residential peaks at 7:30 and 19:00; PAU-2 includes a commercial plateau ranging from 8:00 to 18:00 with sharp transitions; PAU-3 sustains high industrial level around 0.65 p.u. with three-shift variations.

Weekend load reductions differ markedly across PAUs, further substantiating the Non-IID heterogeneity.

The replay validation process strictly adheres to the separation of train/replay rule: none of the parameters are allowed to vary, and there will be no online learning either. At each step, each PAU chooses its action based only on its own local observation, after which the chosen action is fed into the pandapower power flow simulation model to yield the physics result. The evaluation measures cover four categories: the sum total of the cost of operation in the last 30 days, the nodal voltage violations and tie line oscillation

measure, the L2 difference between the sequence of actions and the CPPO optimal policy offline result, and MMD.

These measures are listed in Table 3.

Table 3 summarizes the key metrics under real-world replay conditions.

Table 3. Replay validation results on 30-day real-world data versus simulation baseline

Metric	Simulation Baseline (Full Model)	Real-Data Replay	Relative Deviation
Daily operational cost (CNY)	4,873	5,041	+3.45%
Cumulative 30-day cost (10 ⁴ CNY)	14.62	15.12	+3.42%

Voltage compliance rate (%)	98.4	96.7	-1.7 pp
Tie-line fluctuation rate (%)	4.2	5.8	+1.6 pp
Policy L2 deviation	—	0.137	—
MMD train-replay distribution gap	—	0.082	—

From Table 3, it can be seen that the framework results in an increase in the daily operation cost by 3.45% compared to the baseline scenario simulation, although the percentage of compliance with voltage falls by 1.7% to 96.7%, which is still compliant with the voltage regulation standards of IEEE Std 1547-2018.

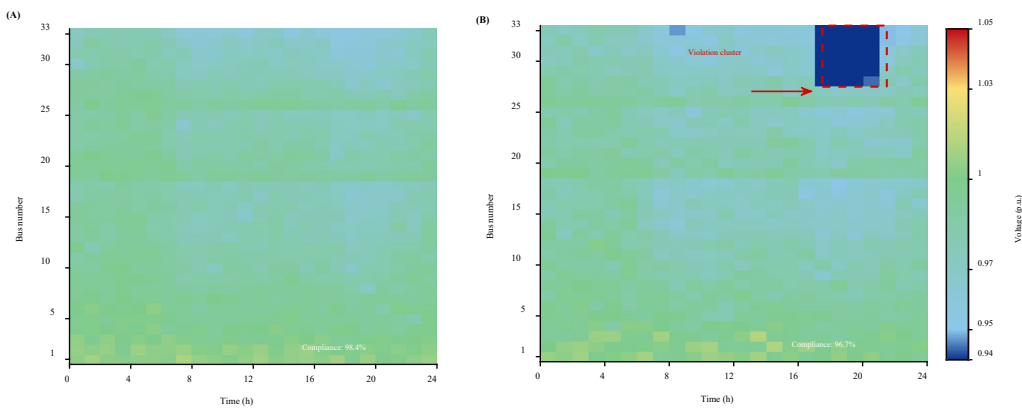


Figure 7. Spatiotemporal nodal voltage heatmaps over 24 hours across all 33 buses

(A) Simulation baseline with 98.4% compliance

(B) Real-world replay with 96.7% compliance, where the violation cluster (red dashed box) is concentrated at PAU-3 remote feeder nodes during evening peak hours

The major component of Fig. 7(A) is the green compliance area with very few violations found throughout the off-peak periods. In Fig. 7(B), there is a compact block of dark blue color for buses 28-33 from 17:00-21:00, where the voltage is less than 0.94 p.u. because of combined factors such as peak demand during evenings and lack of reactive power availability at the end points of feeders. Elsewhere, all buses

are within compliance range, which shows that the 1.7% loss is confined geographically and not system-wide.

The rate of tie-line variation rises from 4.2% to 5.8% with increments taking place in the response periods of two thunderstorms associated with PAU-2 and PAU-3. The deviation of policy L2 equals 0.137 in normalized action space. Action consistency for each PAU is shown in Figure 8.

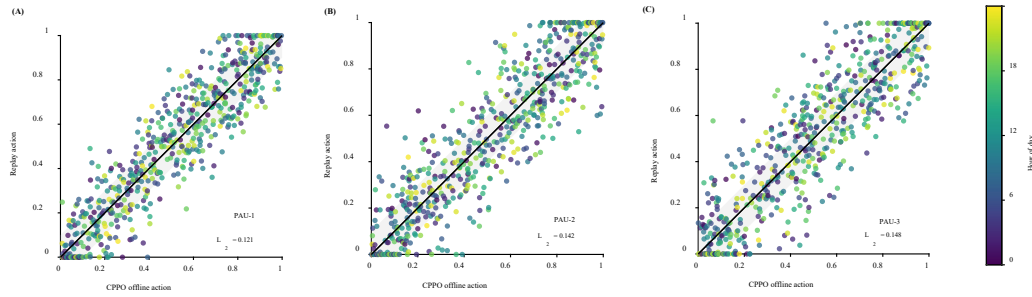


Figure 8. Action consistency scatter plots between Full Model real-data replay and CPPO offline optimal policy
 (A) PAU-1 Residential
 (B) PAU-2 Commercial
 (C) PAU-3 Industrial
 Black diagonal denotes perfect consistency ($y = x$); gray band marks the ± 0.1 deviation region; points are colored by hour of day

As depicted in Figure 8(A), the PAU-1 scatter cloud is compactly clustered in the vicinity of the diagonal region within the ± 0.1 range due to the predictability in residential loads. As observed in Figure 8(B), the cloud associated with PAU-2 demonstrates moderate dispersion especially near the boundary regions of action space that correspond to commercial transitions. The maximum dispersion is observed for PAU-3 (see Figure 8(C)), wherein the outliers in the afternoon range (indicated by yellow-green points) coincide with the thunderstorms experienced on Day 4 and Day 7.

Despite the divergence of MMD being moderate at 0.082 between the training data and the replay data, the policy is still able to generate consistent results using the FedProx proximal constraint. The Full Model also does not have any issues with either policy collapse or control divergence during the whole 30 days of replaying continuously.

5. Discussion

The findings of this study provide insights into the federated proximal policy optimization framework from various perspectives and engage substantively with the existing literature. Regarding convergence properties, the ablation tests show not only how the proposed algorithm behaves numerically but also what problems federated optimization faces with respect to non-IID settings. Without the presence of a proximal term, there would be no convergence in training; it would simply oscillate around the 40th iteration since the cost curve would have a standard deviation 3.2 times higher than that of the entire model. The problem that arises without the presence of a proximal term is due to the training of PPO locally on each PAU and the different distribution of observations, which would cause the local gradients to diverge from the global optimum. During aggregation of such gradients, there is a conflict in the parameter space. Therefore, a proximal term

works as a soft constraint that keeps the local updates constrained within the boundaries of the global model by softly constraining them. This is in keeping with the theory of differential private federated learning, where a constraint of the gradient norm is a must. With heterogeneous data, a direct constraint is needed for converging globally. Both concepts together can give convergence stability and privacy preservation [21].

Privacy-preserving experiments are relevant not only for the results themselves but also due to their implications. Applying two layers of privacy preservation brings the AUC value of the membership inference attack from 0.891 to 0.521, thereby achieving a level of performance close to that of random guessing at a 0.5 value. This phenomenon corresponds to the basic law of information theory, namely, that the degree of inference by an attacker about what training sample was used for a gradient update depends on the mutual information. Gaussian noise perturbation fundamentally disrupts this information channel, while the Rényi differential privacy framework provides a more compact privacy loss combination analysis than the classical (ϵ, δ) -DP through the moment accounting mechanism, allowing the privacy budget consumption after multiple rounds of training to be precisely tracked [22]. Although the overall privacy budget is the same, the AUC for the Fixed DP method is 0.634, while Adaptive DP reduces it to 0.573. This indicates that, in addition to improving utility, time-aware budget allocation also improves privacy protection during critical times. This result is also consistent with the latest trends in MIA attacks in FL, as recent reviews indicate that the timing and intensity of noise injection are more important than the overall noise budget for defense effectiveness [23]. Furthermore, Shokri et al. established a theoretical framework for membership inference attacks against machine learning models, revealing that the overfitting degree of the model to the training samples is a key factor determining the success rate of the attack. The SecAgg method proposed in this paper applies the technique of

masking to avoid the leakage of individual gradients, creating a two-level security defense approach that works complementarily to the addition of DP noise for another attack scenario [24].

In terms of the utility-privacy trade-off relation, the increase in cost caused by the stricter privacy protection parameter $\epsilon=0.5$ is decreased from 11.3% to 8.1% with respect to the time adaptive strategy, a difference of 3.2 percentage points which is of practical importance in an economic aspect of power distribution system planning. This is due to the reason that sensitivity of the gradients is greatly dependent on time periods during the day. At night, when the load is low, this parameter is updated by a much smaller amount than in the evening peak period. Allocating more privacy budget to the most sensitive and high-risk periods provides the greatest protection where information leakage risk is highest, rather than distributing the budget uniformly across the entire day. This directly relates to the discussion in federated learning differential privacy theory on the design of adaptive clipping thresholds. The fixed budget allocation implies uniform gradient sensitivity over time, which clearly does not hold in actual distribution scenarios [21].

Although this study confirms the effectiveness of the proposed framework, certain challenges remain to be noted. The experiments in this study are based on a standard test system with 33 nodes. However, in real-world scenarios with a large number of interconnected multi-microgrid networks, the number of nodes and PAUs is much larger. Hence, the efficiency of the communication and the rate at which the system converges are also a challenge. Another challenge is the scalability of federated learning in large networks, which remains an open challenge for future research [25]. Furthermore, this study focuses on privacy preservation and secure aggregation in the centralized federated learning framework. However, in the context of a decentralized federated learning framework, the challenge is still unresolved, and a systematic approach to classifying the privacy preservation methods is needed [26]. In smart grid applications, current studies have verified that federated learning is both communication-efficient and capable of ensuring privacy in energy forecasting problems. This study can be considered a reference for extending this application to other energy management problems [27]. However, it should be noted that this study only considers steady-state planning. When wind power integration is considered along with permanent magnet synchronous generators and MMC-HVDC in real-world microgrid systems, this introduces a complex set of dynamic stability constraints, representing a promising direction for exploring their coupling with federated optimization.

6. Conclusion

In this study, a privacy-preserving, decentralized planning scheme for a multi-microgrid system is proposed, guided by federated proximal policy optimization. The proposed

scheme leverages the FedProx proximal optimization algorithm, the time-variant differential privacy budget, and the Shamir secret sharing scheme, achieving the balance among convergence stability, cost efficiency, and data confidentiality in Non-IID settings. The experimental results show that the proposed scheme achieves a daily operating cost deviation from the CPPO scheme of no more than 3.0% and reduces the per-round communication load to 9.6 MB (which is about 80% less than the centralized scheme). Furthermore, it reduces the membership inference attack AUC to 0.521, which is just slightly higher than the random guess level, indicating the significance of the dual-layer privacy protection. These findings have important implications for the theoretical understanding and engineering practice of the privacy-aware collaborative planning in large-scale, interconnected multi-microgrid systems under realistic ICT constraints.

Acknowledgements.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] Joshi A, et al. Survey on AI and machine learning techniques for microgrid energy management systems. *IEEE/CAA J Autom Sin.* 2023;10(7):1513-1529.
- [2] Zhao J, et al. Deep reinforcement learning-based model-free on-line dynamic multi-microgrid formation to enhance resilience. *IEEE Trans Smart Grid.* 2022;13(4):2557-2567.
- [3] Liu G, et al. Distributed energy management for networked microgrids embedded modern distribution system using ADMM algorithm. *IEEE Access.* 2023;11:102589-102604.
- [4] Liu G, et al. Microgrid optimal scheduling with chance-constrained islanding capability. *Electr Power Syst Res.* 2017;145:197-206.
- [5] Lee S, Choi DH. Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources. *IEEE Trans Ind Inform.* 2020;18(1):488-497.
- [6] Mukherjee S, et al. Resilient control of networked microgrids using vertical federated reinforcement learning: designs and real-time test-bed validations. *IEEE Trans Smart Grid.* 2024;16(2):1897-1910.
- [7] Cao D, et al. Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices. *J Mod Power Syst Clean Energy.* 2021;9(5):1101-1110.
- [8] Guo C, et al. Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. *Energy.* 2022;238:121873.
- [9] Ye Y, et al. A scalable privacy-preserving multi-agent deep reinforcement learning approach for large-scale peer-to-peer transactive energy trading. *IEEE Trans Smart Grid.* 2021;12(6):5185-5200.
- [10] IShe B, et al. Fusion of model-free reinforcement learning with microgrid control: review and vision. *arXiv preprint arXiv:2206.11398.* 2022.
- [11] Baran ME, Wu FF. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Trans Power Deliv.* 1989;4(2):1401-1407.

- [12] Zimmerman RD, Murillo-Sánchez CE, Thomas RJ. MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans Power Syst.* 2010;26(1):12-19.
- [13] Wilson E, et al. End-use load profiles for the US building stock. Golden (CO): National Renewable Energy Laboratory; 2021.
- [14] Liu H, Wu W. Federated reinforcement learning for decentralized voltage control in distribution networks. *IEEE Trans Smart Grid.* 2022;13(5):3840-3843.
- [15] Chen X, et al. Reinforcement learning for selective key applications in power systems: recent advances and future challenges. *IEEE Trans Smart Grid.* 2022;13(4):2935-2958.
- [16] Hu D, et al. Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization. *IEEE Trans Smart Grid.* 2022;13(6):4873-4886.
- [17] Zhou X, et al. Accelerated voltage regulation in multi-phase distribution networks based on hierarchical distributed algorithm. *IEEE Trans Power Syst.* 2019;35(3):2047-2058.
- [18] Li T, et al. Federated learning: challenges, methods, and future directions. *IEEE Signal Process Mag.* 2020;37(3):50-60.
- [19] Abadi M, et al. Deep learning with differential privacy. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*; 2016.
- [20] Bonawitz K, et al. Practical secure aggregation for privacy-preserving machine learning. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*; 2017.
- [21] Wei K, et al. Federated learning with differential privacy: algorithms and performance analysis. *IEEE Trans Inf Forensics Secur.* 2020;15:3454-3469.
- [22] Mironov I. Rényi differential privacy. In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE; 2017.
- [23] Bai L, et al. Membership inference attacks and defenses in federated learning: a survey. *ACM Comput Surv.* 2024;57(4):1-35.
- [24] Shokri R, et al. Membership inference attacks against machine learning models. In: *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE; 2017.
- [25] Li T, et al. Federated optimization in heterogeneous networks. *Proc Mach Learn Syst.* 2020;2:429-450.
- [26] Yin X, Zhu Y, Hu J. A comprehensive survey of privacy-preserving federated learning: a taxonomy, review, and future directions. *ACM Comput Surv.* 2021;54(6):1-36.
- [27] Badr MM, et al. Privacy-preserving and communication-efficient energy prediction scheme based on federated learning for smart grids. *IEEE Internet Things J.* 2023;10(9):7719-7736.