湘潭大学 自动化与电子信息学院
XIANGTAN UNIVERSITY

Power System-Aware Adversarial Reinforcement Learning Attacks and Defenses on Single and Multi-Agent Dispatch across Continuous / Discrete Domains

面向电力系统的对抗强化学习：连续/离散域下单智能体与多智能体调度的攻击与防御研究

Dr. SU YI (苏 译)

Email: suyi2018@xtu.edu.cn

2025.04.28

Biography

Yi,Su （苏译）, Ph.D，USM博士后，湘潭大学（双一流）助理教授，硕士生导师。

Education Background：

2009.09-2013.07: Wuhan University of Technology (211) 武汉理工大学	B.E. in Electrical Engineering and Automation 电气工程及其自动化专业
2013.09-2016.07: Hunan University (985) 湖南大学	M.E. in Electrical Engineering 电气工程专业（推免）
2020.10-2024.03: Universiti Sains Malaysia (2024"QS" :137) 马来西亚理科大学	Ph.D. in Power Systems and Energy Conversion 电力系统及能源转换

Work Experience：

2016.09-2020.09:	Zhongshan Power Supply Bureau, System Department 广东电网中山供电局，系统部	Automation and Operations 调度自动化
2024.04-Present:	Universiti Sains Malaysia (USM) 马来西亚理科大学	Postdoctoral Fellow 博士后
2024.03-Presen:	Xiangtan University 湘潭大学（双一流）	Assistant Professor / Master's Supervisor 助理教授，硕士生导师

Academic Achievements and Awards:

- Published over 10 papers as the first author/corresponding author, including 5 JCR Q1 papers and 2 ESI Highly Cited Papers.
  以第一作者/唯一通讯作者发表SCI/EI论文十余篇，其中JCR一区论文5篇，ESI高被引论文2篇。
- Participated in 7 major technology and engineering projects, including: Malaysia's Zero-Carbon Island, etc.
  完成7项科技项目和工程项目，包括：马来西亚零碳岛屿（国家级示范工程）;广东电网数字化变电站及配电自动化改造
- Awarded USM Full Scholarship(2021-2024), National Graduate Scholarship (2015), Second Prize of Hunan Electric Power Science and Technology Award (2015),etc.

Team Profile

(a) 马来西亚国家级零碳岛屿示范工程

厂房屋顶及车棚分布式光伏发电

装机容量4.54MWp，年发电量458万kWh，园区年用电量330万kWh，园区整体实现碳中和运营。

园区储能电站

装机容量100kW/300kWh，可满足园区核心负荷用电1.5至2小时，并与市电、光伏发电、柴油发电机构成园区微网，可实现离并网运行的切换。

园区空气源热泵热水系统

园区电动车及换电系统

园区能源监控中心及需求侧管理平台

(b)湘潭市九华园区光储充示范项目

◆ 马来西亚理科大学(1969-)是马来西亚教育部唯一指定的APEX（迈向卓越计划）大学，**2024年全球QS排名第137位**，同时也是国内“一带一路”高校联盟的重要成员。欢迎一起联合申报科技部重点研发项目-政府间国际科技创新合作专项。

◆ 湘潭大学（1958-）是全国16所文理工综合性重点大学之一，入选**国家“双一流”建设高校**，拥有国家应用数学中心（国家级），智能计算与信息处理教育部重点实验室，多能协同控制技术等多个相关科研平台。所在的智能电网与人工智能团队长期致力于智能电网与人工智能领域的交叉研究，近5年来承担省部级及以上科研课题30余项，科研经费超过1000万元。

Journal papers (Part)

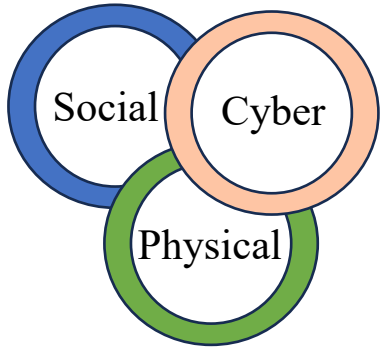
1. **Su, Y.**, *et al*., "Attack and Defense Against Deep Reinforcement Learning-Based Agent for Discrete Domains: Application in Distribution Network Reconfiguration, " ***IEEE Trans. Autom. Sci. Eng.***(Major revision)
2. Z. Dai, **Su, Y.,*** *et al*., " Adversarial Attacks and Defense for Multi-Agent Deep Reinforcement Learning-Based Optimal Dispatching in an Active Distribution Network , " ***IEEE Trans. Syst. Man Cybern. Syst.*** (Major revision)
3. **Su, Y.,** *et al.,* " Optimal Dispatching of Active Distribution Networks Considering the Adversarial Attacks Using the Deep Reinforcement Learning, " ***IEEE Trans. Smart Grid.*** （Under review）
4. Z.Dai, **Su, Y.,*** *et al*., " Auto-shielded Robust Reinforcement Learning for Electric Vehicle Charging Scheduling Under Multi-Criteria Adversarial Attacks "***IEEE Trans. Transp. Electrific***. （Under review）
5. **Su, Y.**, *et al*., " Optimal dispatch for AC/DC Hybrid Distribution Systems with Electric Vehicles: Application of Cloud-Edge-Device Cooperation," ***IEEE Trans. Intell. Transp. Syst.,***2024.
6. **Su, Y.,** *et al.,* " Short-term Transmission Capacity Prediction of Hybrid Renewable Energy Systems Considering Dynamic Line Rating Based on Data-driven Model, " ***IEEE Trans. Ind. Appl.,*** 2025.
7. **Su, Y.** & Teh, J. " Two-Stage Optimal Dispatching of AC/DC Hybrid Active Distribution Systems Considering Network Flexibility," ***J. Mod. Power Syst. Clean Energy.,*** 2023.
8. **Su, Y.**, *et al*., " A Two-Layer Framework for Mitigating the Congestion of Urban Power Grids Based on the Flexible Topology with Dynamic Thermal Rating, " ***Prot. Control Mod. Power Syst.,*** 2024.
9. **Su, Y.,** *et al*., " Two-layer Optimal Dispatch Framework of Active Distribution Networks with Hybrid Energy Storage via Deep Reinforcement Learning and Real-time Feedback Dispatch," ***J. Energy Storage.,*** 2025.
10. **Su, Y.** & Teh, J. *" Hierarchical and Distributed Energy Management Framework for AC/DC Hybrid Distribution Systems with Massive Dispatchable Resources, " ***Electr. Power Syst. Res.,***2023.
11. **苏译**,等. 考虑不确定性扰动的含海量资源配电网弹性研究综述[J]. ***电力系统自动化***.,2022.

Contents

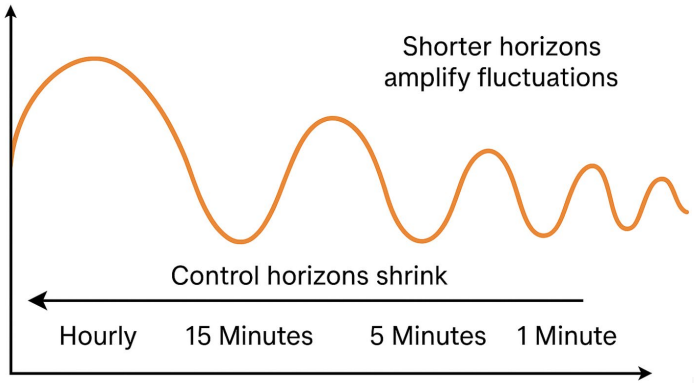
Why DRL for the ADNs?

Modern Active Distribution Network Landscape

10^5 + Dispatchable Assets



Cyber-Physical-Social Energy System (CPSES)



Real-time Control (1-5 min)

- 10^5 + controllable nodes across storage, flexible load, EVs and distributed energy resources (DERs) in a modern ADNs.
  一个大型城市的新型主动配电网可调度资源超过10^5
- The cyber, physical, and social layers interact bidirectionally, forming a CPSES.
  信息、物理与社会三层面相互影响，构成信息-物理-社会耦合的能源系统
- Decision space becomes more dimensional and more constrained — a hallmark of **MILP**-class problems.
  决策空间维度和约束显著上升，呈现出 MILP 问题的典型特征
- Uncertainties propagate and amplify across the cyber, physical, and social layers, demanding faster real-time dispatch, such as 1-5 min.
  不确定性在CPSES传递并被放大，对实时调度提出了要求，如1-5分钟尺度
- Large-scale MILP formulation of dispatch—formally NP-hard—poses a fundamental tension with real-time dispatch.
  调度的大规模 MILP 本质上是 NP难题，与实时调度形成根本冲突

Why DRL for the ADNs?

From Conventional Optimization to DRL Paradigm

Conventional Optimization

- MILP scale exponentially with decision accounts.

MILP随决策变量指数级增加

- Dispatching results depend on the forecasting accuracy.

调度结果依赖预测精度

- Static modelling is no longer sustainable

静态模型难以适应动态环境

Computation-heavy, Static & Inflexible

Why Transition

PJM, CAISO, SGCC;
IEEE P2801 draft
(AI Dispatch)

DRL Paradigm

- DRL scales extend via policy networks, which grows linearly

DRL不直接求解全局MILP，推理阶段的复杂度随网络线性增长

- DRL makes real-time decisions through environment interaction.

DRL通过与环境交互实时决策，不依赖预测结果

- DRL dynamically fine-tunes parameters online.

DRL在线动态微调参数适应系统动态演化

Model-free, Adaptive & Scalable

Emergent Vulnerabilities in DRL-driven ADNs Dispatch

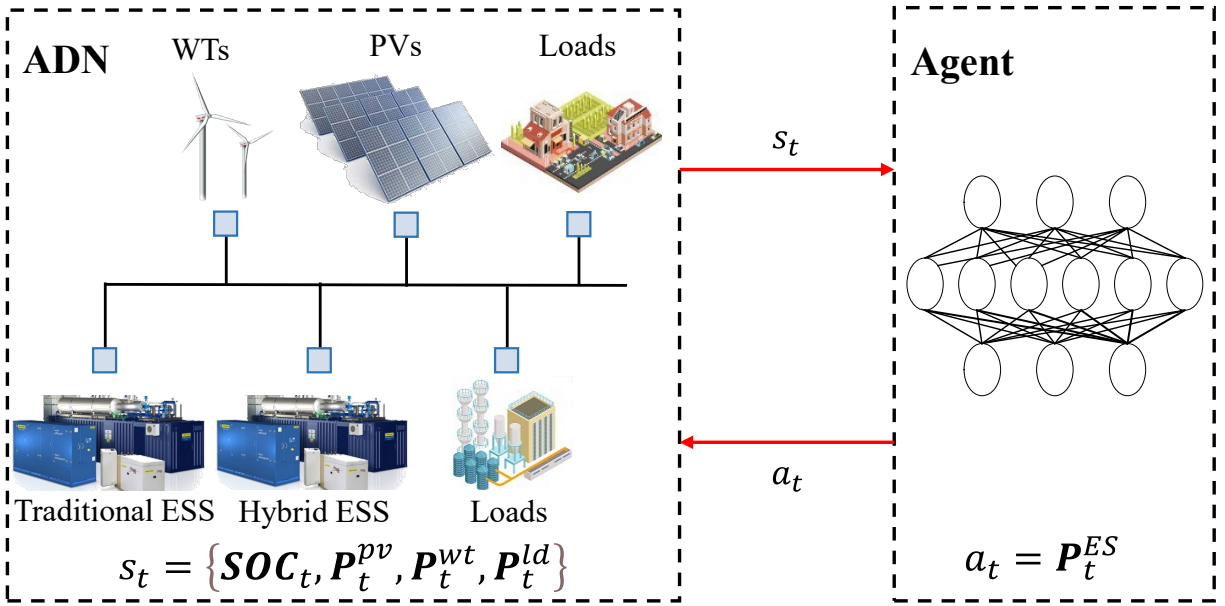
DRL-driven Dispatch Process in ADNs

- States(状态):$s_t = \{\boldsymbol{SOC}_t, \boldsymbol{P}_t^{pv}, \boldsymbol{P}_t^{wt}, \boldsymbol{P}_t^{ld}\}$;
- Actions(动作): $a_t = \boldsymbol{P}_t^{ES}$
- Rewards(奖励): The agent receives an immediate reward r_t when the stage transitions from s_t to s_{t+1} with action a_t. Namely, the r_t is the accumulation of the objective function over the time steps, including the objective function and the discharge/charge of ESS within a reasonable range after one day.

$$r_t = \begin{cases} -F(a_t) & t \neq T \\ -F(a_t) - G & t = T \end{cases}$$

- State transition(状态更新): state is considered as measured values after the action a_t finished.
- Objective and discount factor (优化): The objective of each agent is to maximize the cumulative rewards:

$$R = \sum_{t=1}^{T} \gamma^t r_t.$$

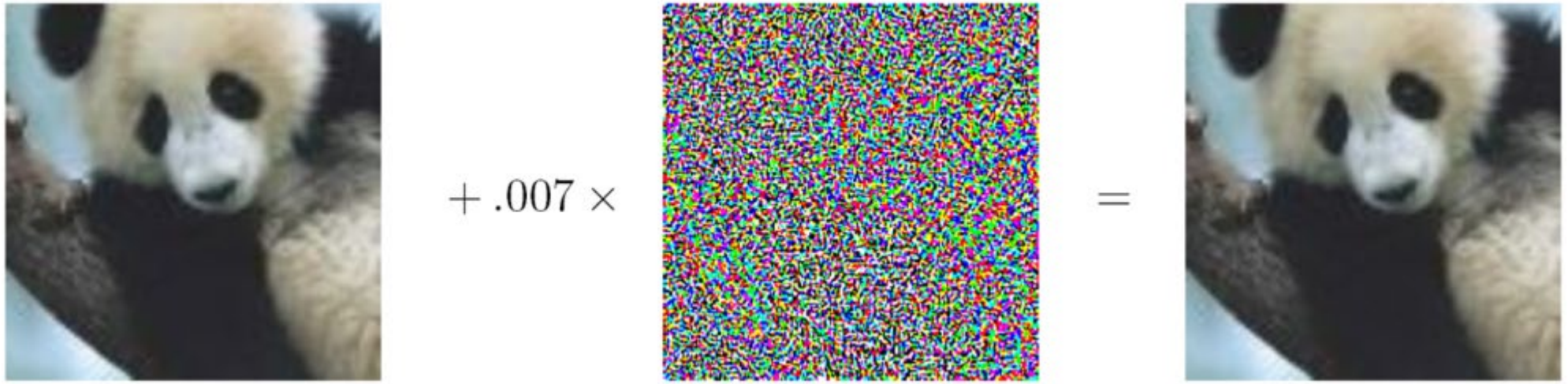


Thus, the dispatch of ADNs can be defined as finding an optimal strategy π to maximize the expectation of cumulative rewards R: max $\mathbb{E}_\pi[\sum_{t=1}^{T} \gamma^t r_t \,|s_t, a_t]$

最优决策

Emergent Vulnerabilities in DRL-driven ADNs Dispatch

New Vulnerabilities of DRL: Minor disturbance, Major Error



Panda **Slight** disguised data **Gibbon**

- **Slight Data Injection**: Can lead to entirely different classification results.
  轻微扰动可能使得分类失败，即产生错误的决策

First-order Taylor expansion term

disturbance

$$f(\tilde{x}) = f(x+\epsilon) \approx f(x) + \nabla f(x) \cdot \epsilon + o(\|\epsilon\|)$$

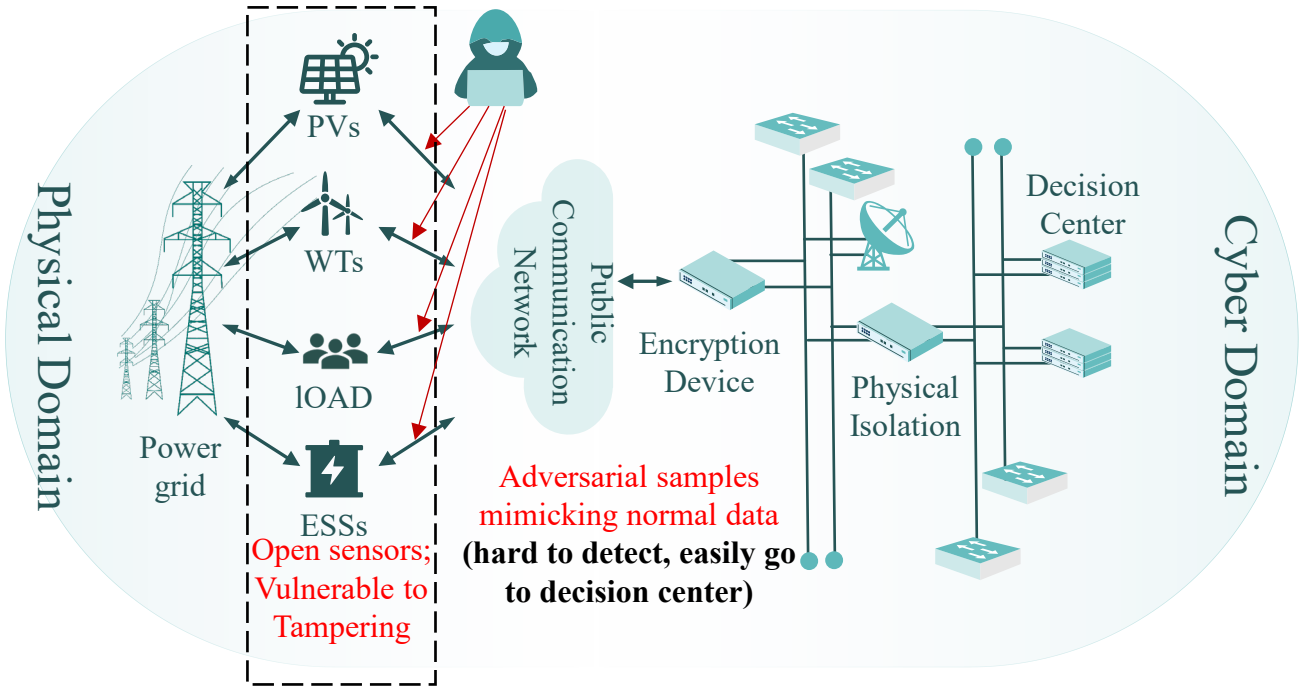
local gradient sensitivity

In deep neural networks, the chain rule across multiple layers further amplifies the initial perturbation, ultimately leading to significant decision deviations. 深度神经网络中，上述波动通过多层链式乘积效应进一步放大

Thus, in DRL-driven distribution network dispatch, **injecting adversarial data that resemble normal data** can cause completely erroneous dispatch decisions. 注入与正常数据相似的恶意数据，可能导致配电网DRL智能体错误

Emergent Vulnerabilities in DRL-driven ADNs Dispatch

Attack Feasibility in DRL-driven Active Distribution Networks



Stealthy adversarial attacks can infiltrate DRL-driven ADNs via the "open-edge sensors → firewall → dispatch center" pathway, bypassing conventional security layers.

DRL调度系统可通过“开放终端传感器→防火墙→调度中心”路径受到隐蔽对抗攻击影响

◆ Open and Vulnerable Distributed Sensors

开放且脆弱的分布式资源终端传感器

Sensors deployed at distributed resources (PVs, WTs, ESSs) are poorly protected, making them susceptible to tampering and malicious data injection.

◆ Bypassing Detection through Stealthy Perturbations

轻微扰动可绕过安全防护装置

Large deviations are blocked by firewalls,but small, normal-like adversarial modifications can bypass perimeter defenses and infiltrate the control center.

◆ DRL is highly sensitive to minor disturbance.

DRL智能体对微小扰动高度易感

In traditional dispatch, small disturbances are constrained by physical modeling; whereas in DRL, local gradient sensitivity amplifies perturbations.

Adversarial Attack Strategies & Results for ADNs

3-1 Single-Agent Adversarial Attacks –Using Local Gradient Sensitivity

单智能体连续域对抗攻击以验证DRL驱动的ADNs局部梯度敏感性

3-2 Multi-Agent Adversarial Attacks –Considering Interaction Among Agents

多智能体连续域对抗攻击利用合作交互导致跨智能体更大范围的优化失败

3-3 Topology Reconfiguration Attacks –Flipping in Discrete Domain

不显著改变奖励值同时突破物理操作限制（频繁拓扑变化）

Adversarial Attack Strategies & Results for ADNs

Single-Agent Adversarial Attacks--Methodology

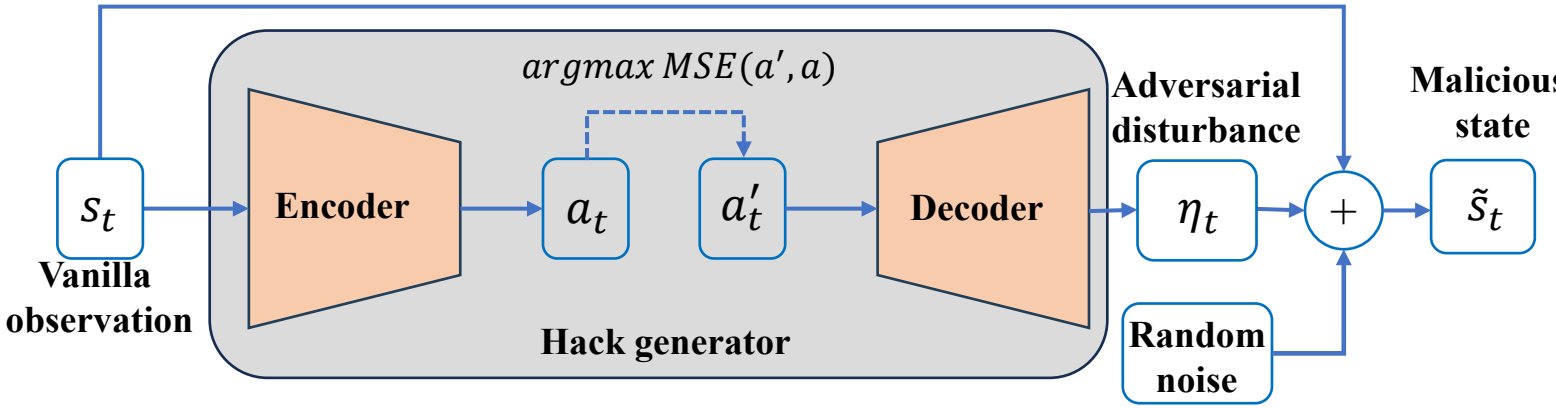
Proposes the Model INterference via MALicious noise (MINMAL) method, which maximizes the deviation between the model's output actions and the correct actions. 基于动作偏差最大化，在允许波动内反向求解攻击

1. Find an adversarial action a' that **maximizes the loss function**:

$$\pi(s') = \max_{\|\delta\| \le \varepsilon} \mathcal{L}(s+\delta, a)\Big|_{\varsigma}$$

2. To navigate away from local minima, introduce a random perturbation η to the original sample s.

$$\max_{\|\delta\| \le \varepsilon} \mathcal{L}(s+\eta+\delta, a)\Big|_{\varsigma}$$



3. Employ a multi-step gradient ascent method to iteratively update the perturbation δ

$$\delta_{t+1} = \delta_t + \alpha \cdot \text{sign}\left(\nabla_s \mathcal{L}\left(s+\eta+\delta_t, a\right)\Big|_{\varsigma}\right)$$

4. Projection operation is applied to the updated perturbation to ensure the generated adversarial samples stay within the allowable perturbation range

$$\delta_{t+1} = \min\left(1, \frac{\varepsilon}{\|\delta_t\|}\right) \cdot \delta_t$$

5. By iteratively adding small perturbations and performing projection at each step to **make the generated adversarial perturbations close to the original state**.

Adversarial Attack Strategies & Results for ADNs

Single-Agent Attack Results – All Sensors

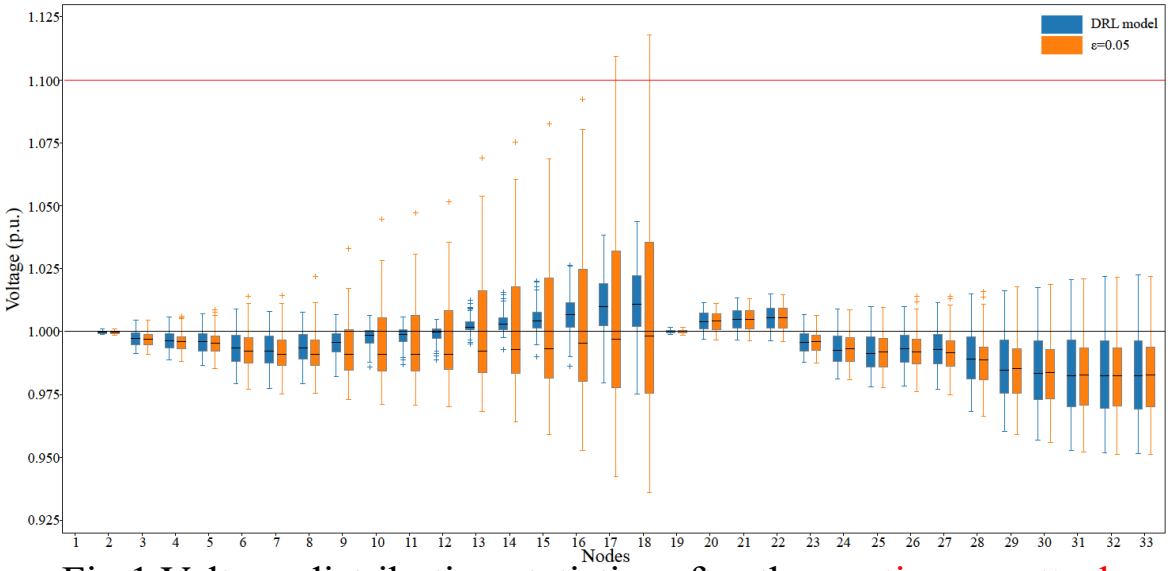


Fig.1 Voltage distribution statistics after the continuous attacks

TABLE I Voltage under different attack amplitudes

Vol (p.u.)	Un-attacked	MINMAL			
		0.05	0.1	0.2	0.4
Min	**0.952**	0.936	0.860	0.795	0.775
Max	**1.044**	1.118	1.148	1.125	1.105

- Even a small attack of ε=0.05, cause the DRL-based ADN to produce overvoltage results.

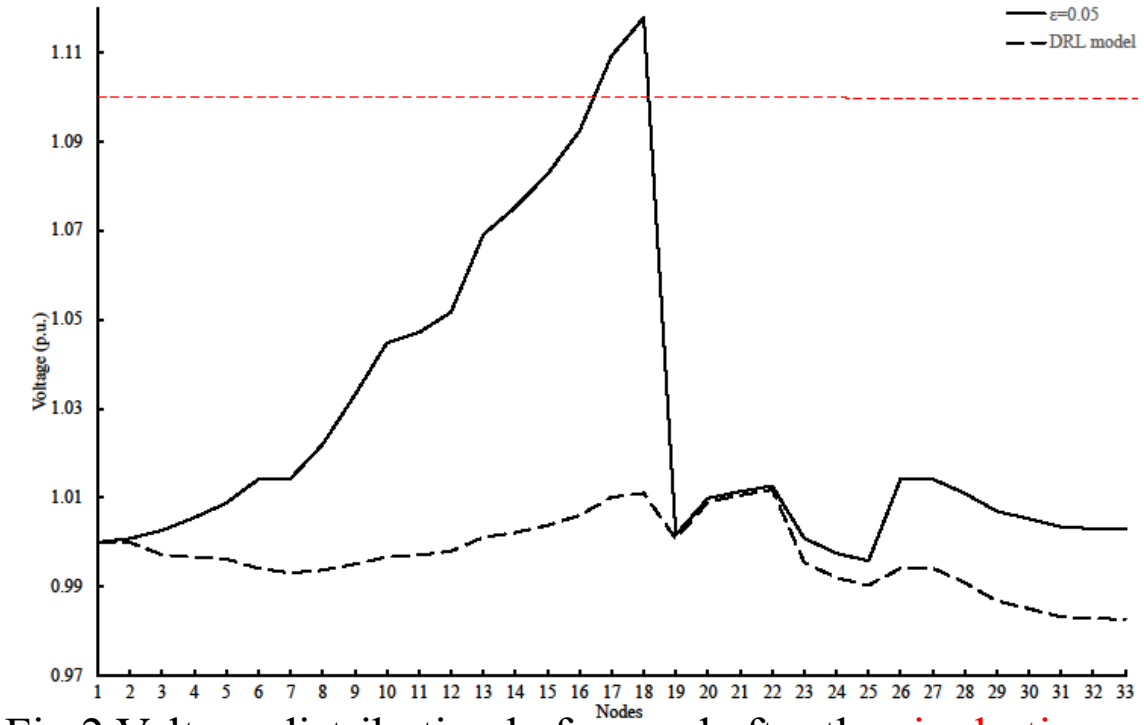


Fig.2 Voltage distribution before and after the single-time attack

- A single-time adversarial attack injected at a critical moment can trigger overvoltage conditions, compromising grid stability.

Adversarial Attack Strategies & Results for ADNs

Single-Agent Attack Results – Section of Sensors

TABLE II Comparisons in Different Single-dimensional Attacks

Vol (p.u.)	Un-attacked	ESS		PV		WT		Load	
		$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.05$	$\varepsilon = 0.1$
Min	**0.952**	0.952	0.952	0.938	0.936	0.935	0.934	0.937	0.935
Max	**1.044**	1.044	1.043	1.108	1.113	1.112	1.115	1.096	1.109

- Even perturbing specific sensor types—an attack scenario that **is easier to implement**—still causes major decision deviations, with WTs being the most vulnerable; SoC of the ESSs show negligible impact.

Single-Agent Attack Results – Single sensor

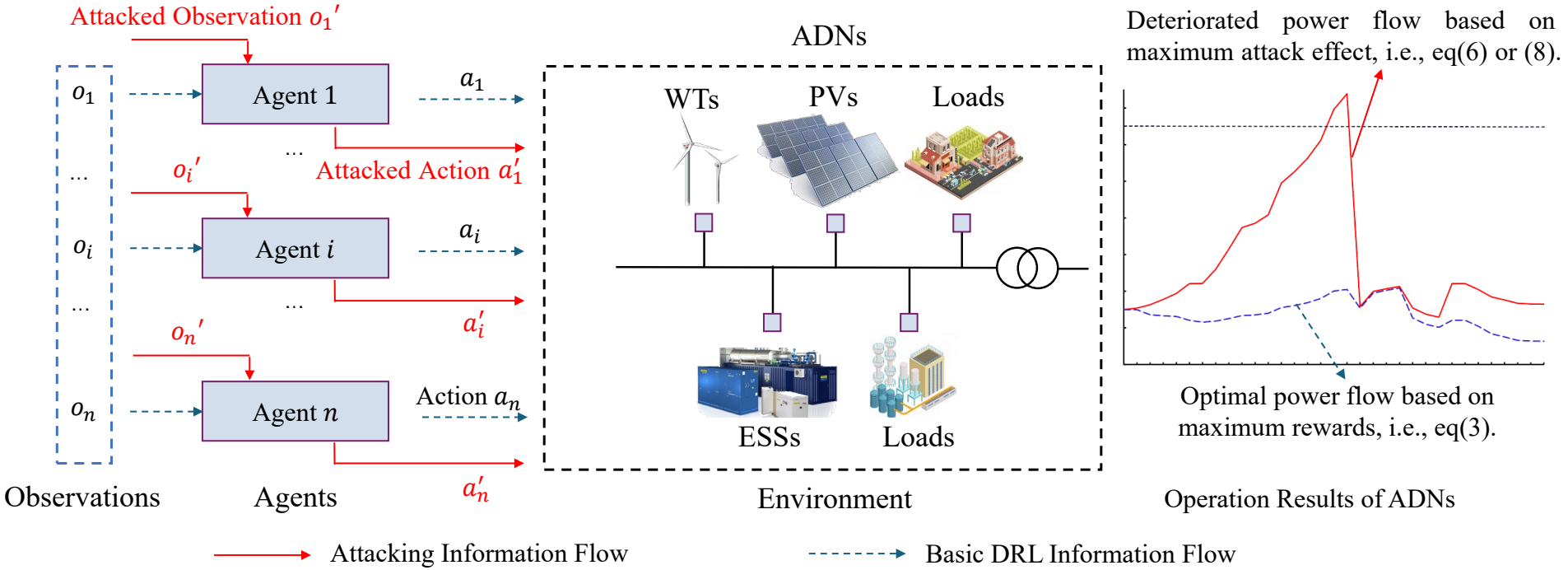
Vol (p.u.)	Un-attacked	ESS	PV	WT	Load
		$\varepsilon = 0.05$	$\varepsilon = 0.05$(Node 27)	$\varepsilon = 0.05$(Node 32)	$\varepsilon = 0.05$(Node 18)
Min	**0.952**	0.952	0.935	0.935	0.936
Max	**1.044**	1.044	1.104	1.108	1.096

- Single-sensor attacks, which are the **easiest to implement**, can induce significant decision deviations by injecting small perturbations at the most vulnerable node in the grid.

Compared with **random noise attacks**, and **gradient-based attacks, using different DRL models**. Results show that the proposed attack is useful and powerful.

Adversarial Attack Strategies & Results for ADNs

Multi-Agent Adversarial Attacks --Methodology



- This attack can be propagated and amplified by the interaction between all agents(对每个智能体施加梯度优化+稀疏性引导) :

$$\delta_{i,t}^{k+1} = \delta_{i,t}^{k} + \alpha_{\mathcal{L}} \nabla_{\delta_{i,t}} \mathcal{L}_{i,t}^{att} + \alpha_d \frac{1}{|\mathcal{O}_i|} \sum_{c=1}^{|\mathcal{O}_i|} (\delta_{i,t,c} - \delta_{i,t}^{k})$$

Gradient Attack **Sparse Diffusion**

- If the output disturbance is in the same direction of all agents, the accumulation of multiple disturbances will result in greater fluctuation (每个时间步对每个目标智能体逐步施加精确扰动，达到所有智能体统一动作进行最大破坏):

$$\delta_{i,t} \sim \mathcal{P} = softmax\left(\mathcal{L}_{i,t}^{att} \Big|_{a_{i,t}' = \mu_{\pi_i}(\boldsymbol{o}_{i,t} + \boldsymbol{\delta}_{i,t,c})}, \forall\, \boldsymbol{\delta}_{i,t,c} \in \mathcal{O}_i \right)$$

Adversarial Attack Strategies & Results for ADNs

Multi-Agent Adversarial Attacks –Part of Results

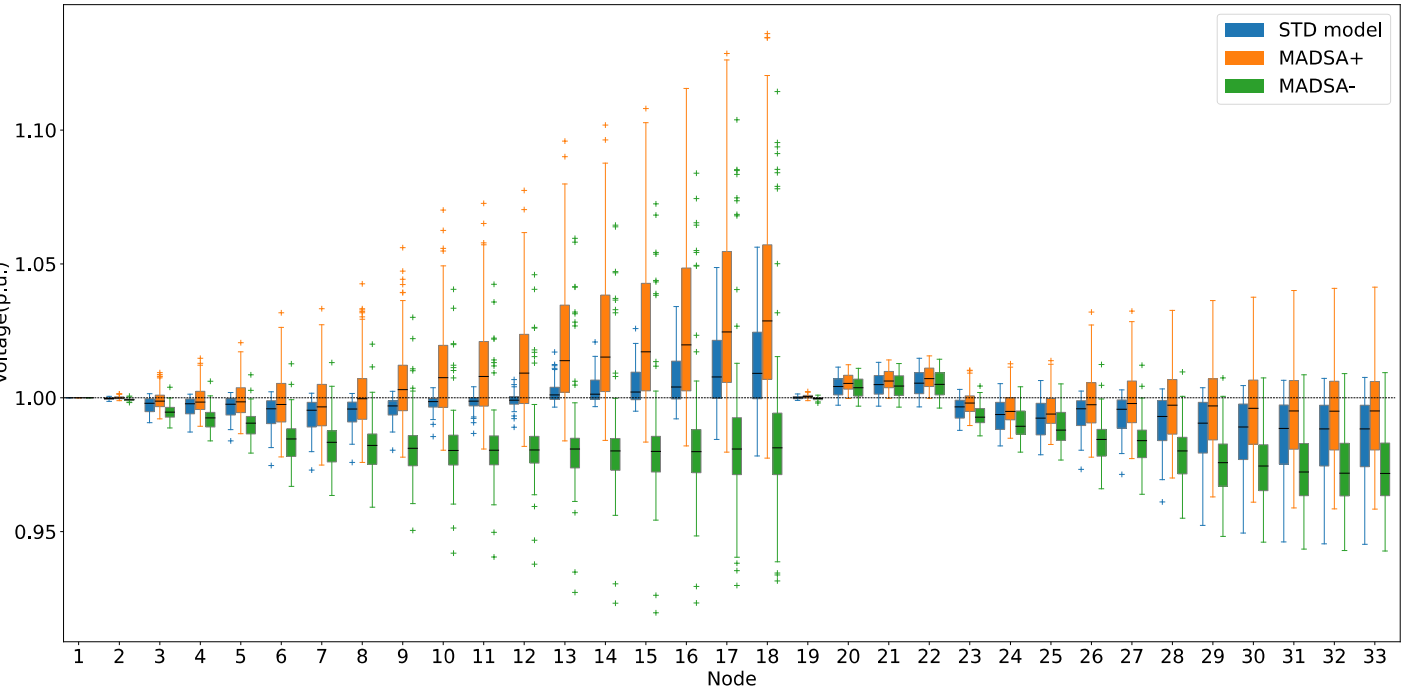


TABLE I

RESULTS UNDER DIFFERENT ATTACKS AGAINST FULL-AGENT

	F^{obj}		
	0.05	0.1	0.2
Std	9.405		
FGSM	13.938	19.839	31.123
Noise	10.483	10.936	13.592
MADSA^{s}	17.623	26.382	39.678
MADSA^{+}	23.459	40.819	44.315
MADSA^{-}	22.936	39.963	54.785

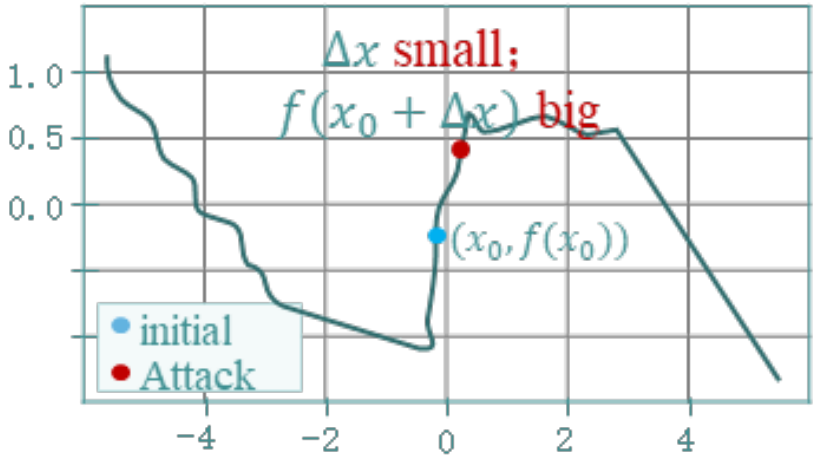
Fig. 1. Voltage statistics when MADSA selects different attack directions.

- Voltage deviations become significant, shifting either upward or downward depending on the attack direction.
  所述攻击方法通过诱导智能体朝同一方向偏移，放大扰动效应，加速系统退化。

- Proposed method achieves significantly stronger adversarial effects than conventional attacks, as shown in the rewards.
  从奖励值变化可以看出，所提出的攻击方法能够显著劣化系统性能，攻击效果更加突出。

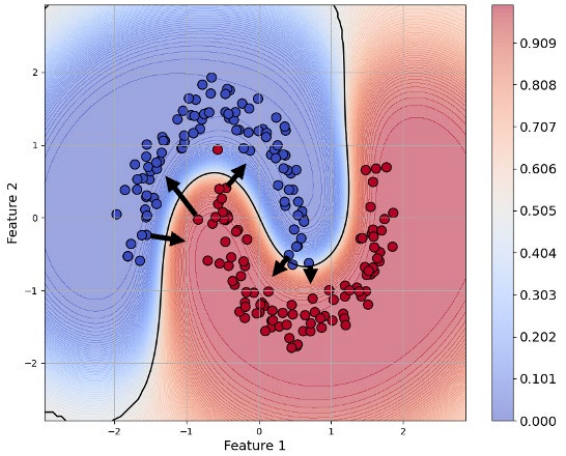
Adversarial Attack Strategies & Results for ADNs

Topology Reconfiguration Attack– Methodology applicable to discrete domains



Continuous domain：local gradient sensitivity

连续域：局部梯度敏感性



Discrete domain：action flip

离散域：动作翻转但奖励值变化不大

- In continuous domains, small perturbations can cause large continuous output deviations.
  梯度敏感性导致奖励值大幅变化
- In discrete domains, small perturbations can directly flip decision categories.
  奖励值波动不大但是决策完全翻转

Eq	Function	Description
$\mathrm{KL}(p\|\|q)=\sum_i p_i \log \frac{p_i}{q_i}$	Decision Bias Inducer	Measures distribution shift before and after attack
$\mathrm{CW}(x,t)=\max(\max_{i\neq c}(f(x)_i - f(x)_c + \kappa, 0)$	Adversarial Strength Inducer	Increases confidence in incorrect actions
$\mathcal{L}_{\text{total}} = \sum_{l=1}^{N} (\frac{1}{w_l} \cdot \frac{1}{\sum_{k=1}^{N} 1/w_k}) \mathcal{L}_l$	Adaptive Weighting	Dynamically balances multiple attack objectives

APGA-P

Non-directional attack;
Increased system loss;
Dramatic changes in reward

APGA-F

Directed attack;
Frequent changes in topology;
Small changes in reward

Adversarial Attack Strategies & Results for ADNs

Topology Reconfiguration Attack– Results

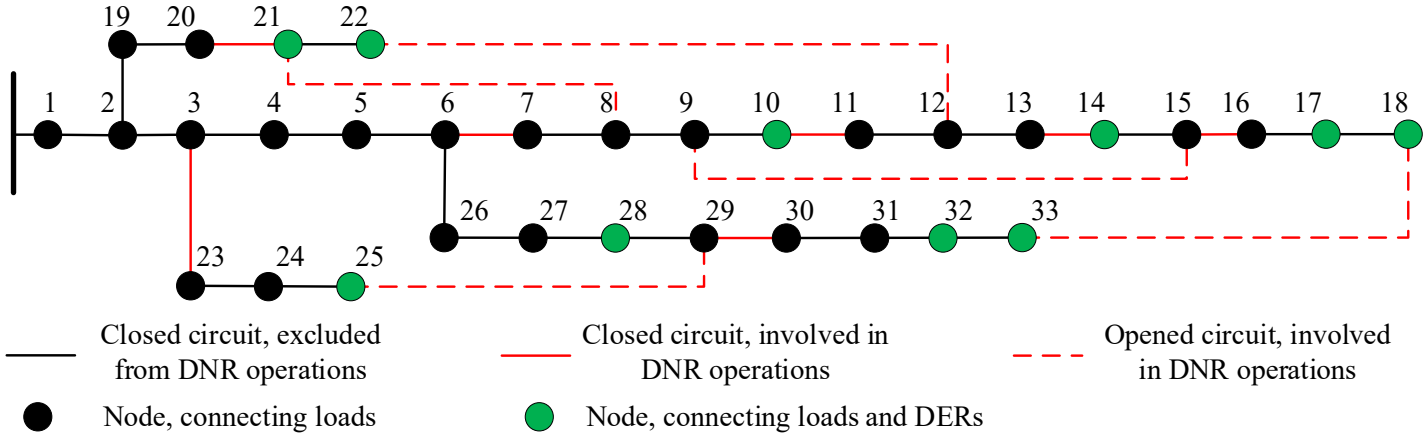
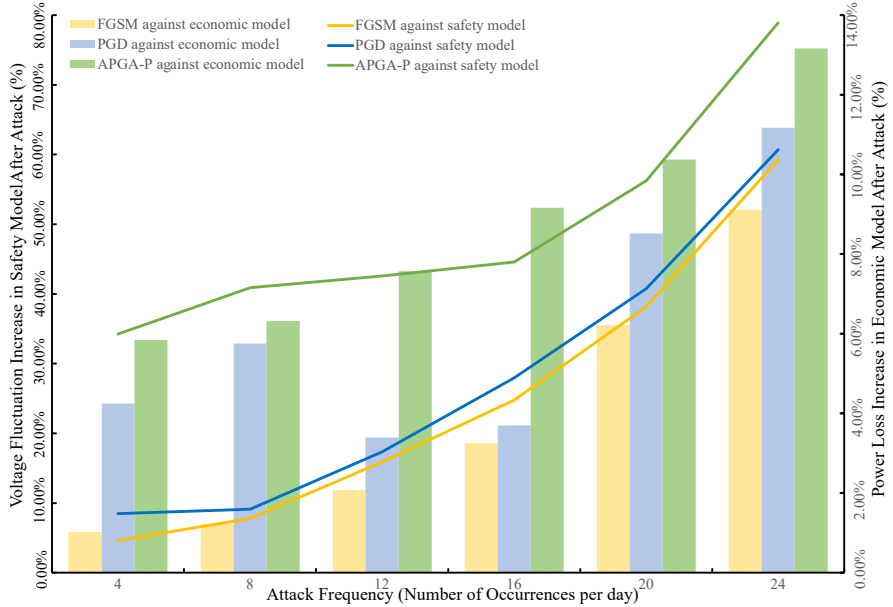
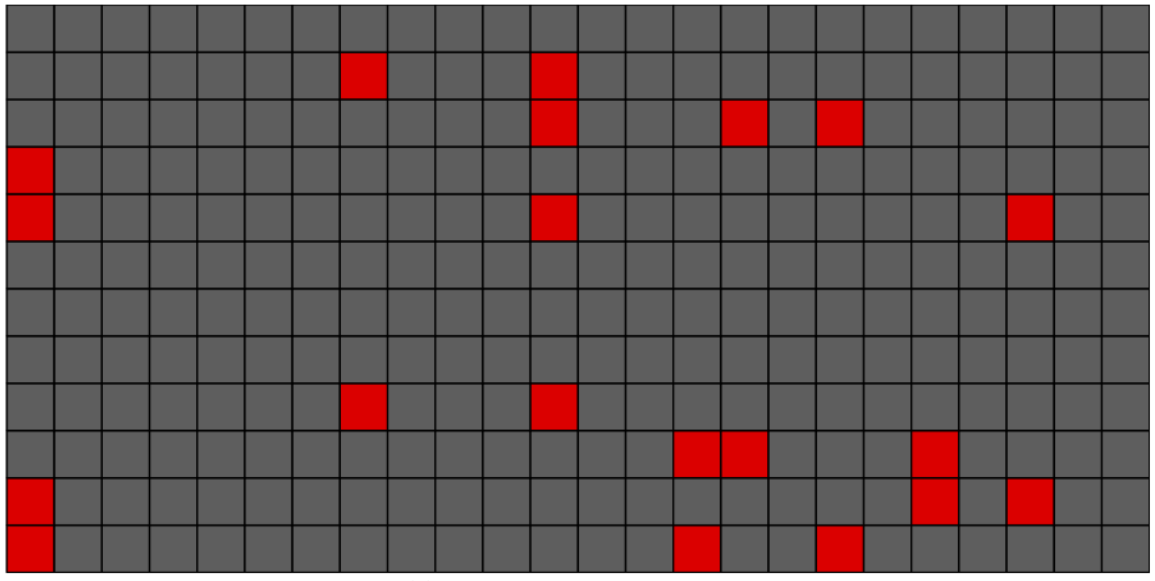
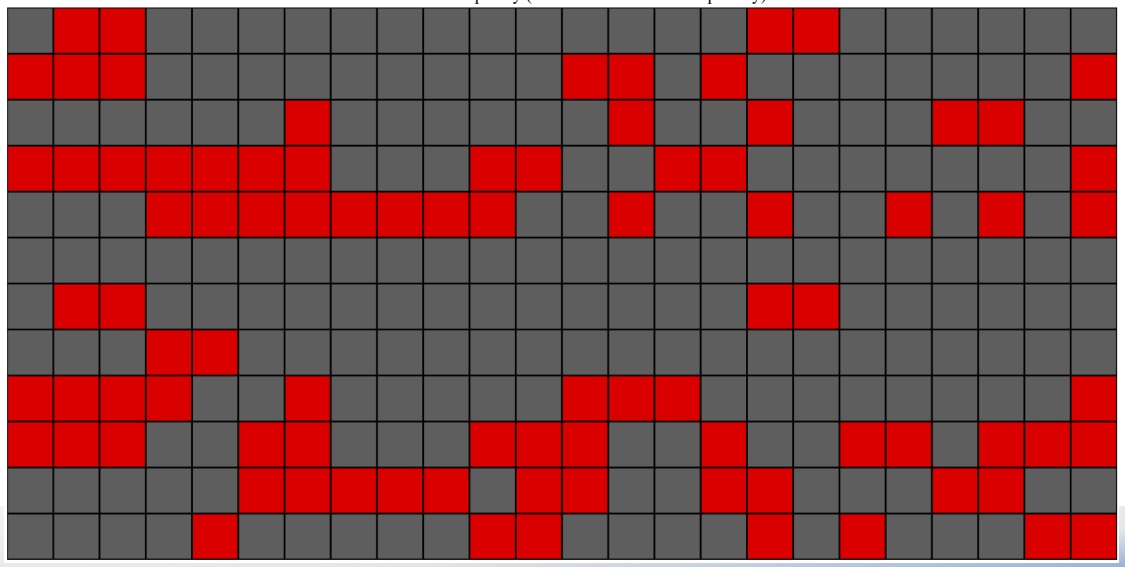


Fig.1 Impact of varying attack frequencies on system degradation





STD



ATT

Defense Strategies & Results for ADNs

4-1 Adversarial Training: Basic Defense Foundation

基本对抗样本训练

4-2 Curriculum Adversarial Training (CAT): Progressive Strengthening

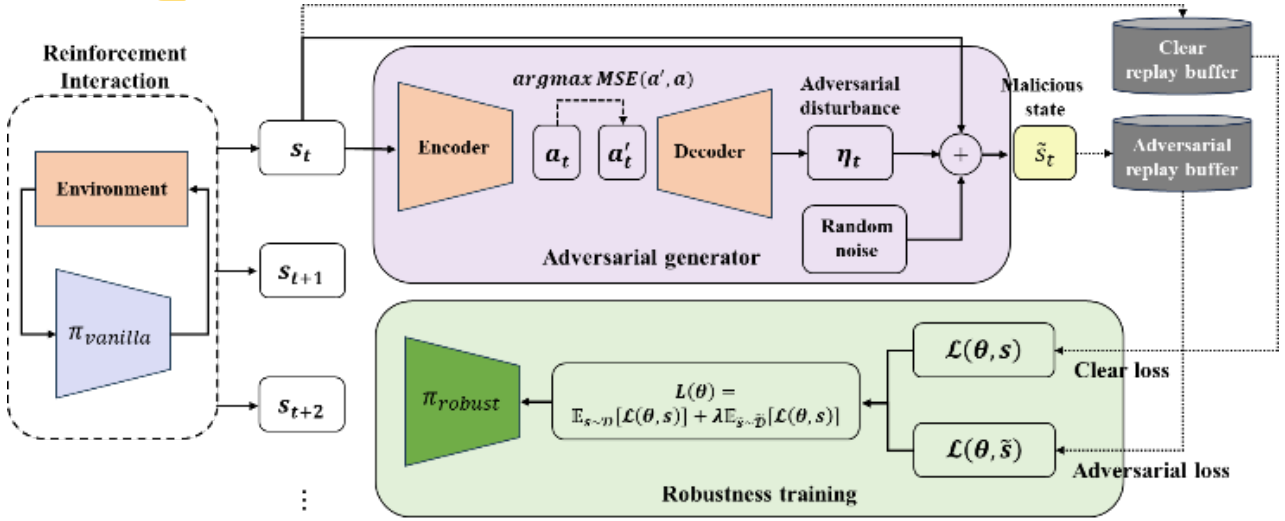
逐步增强对抗样本的课程式对抗训练，提升训练稳定性

4-3 Model Enhancement: Accelerating Robust Learning

改进深度模型结构，加速鲁棒收敛

Defense Strategies & Results for ADNs

Adversarial Training: Basic Defense Foundation—Methodology



- Dual-buffer strategy to avoid overfitting of one attack.
  双缓冲策略避免过拟合某种攻击

- Adversarial training is dynamic and self-adaptive, responding to different attack types during training.
  对抗样本需要根据对抗攻击动态调整和自适应

Adversarial Training: Basic Defense Foundation—Results

TABLE I
Defense Results Against Continuous Attacks

	Mi voltage (p.u.)	Max voltage (p.u.)	Power loss (MWh)	$\sum_T F^{ob}$
Basic	0.952	1.044	1.39	10.88
Un-defensed	0.936	1.118	1.98	16.98
Defensed	0.951	1.086	1.62	14.09

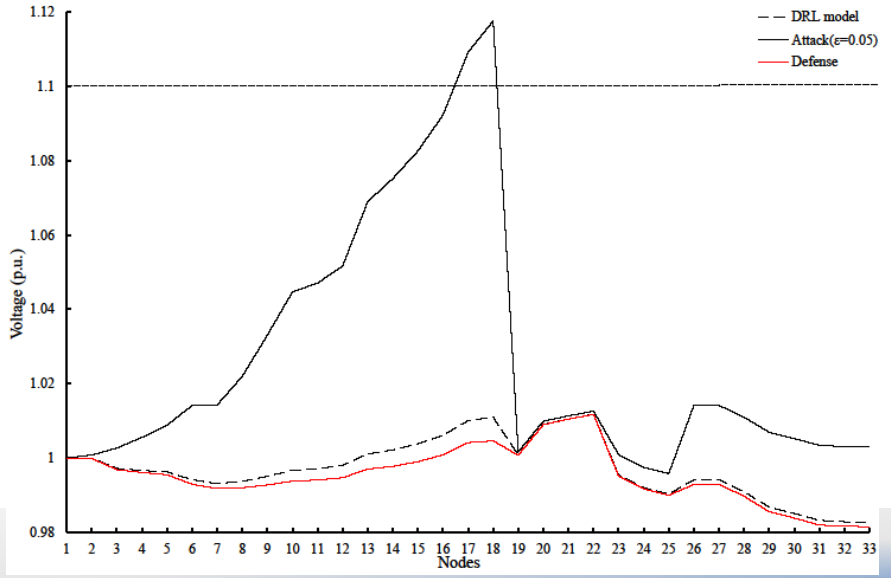
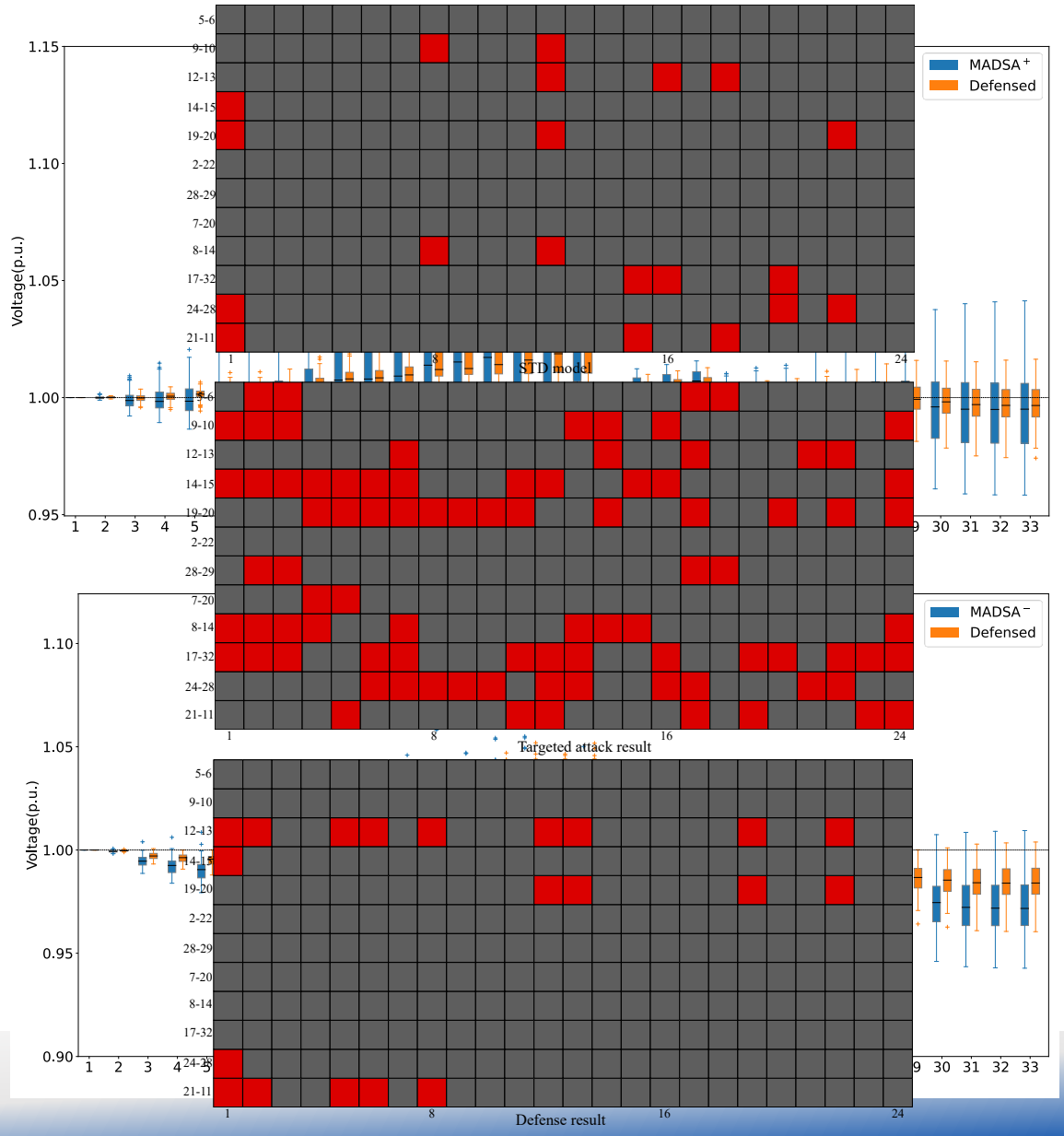


Fig. 1. Voltage before and after the defense.

Defense Strategies & Results for ADNs

Curriculum Adversarial Training (CAT)

- Initial Task (Easy Adversarial Samples)
- Increase Difficulty (Moderate Adversarial Samples)
- Introduce Multiple Attack Modes (Advanced Adversarial Samples)
- Increase Sample Concentration and Attack Magnitude
- Comprehensive Training and Evaluation
- Final Full-Range Adversarial Training



Defense Strategies & Results for ADNs

Model Enhancement: Accelerating Robust Learning

Defensive Strategy Based on Gradient Leveling Regularization (GLR) and a New Neural Network

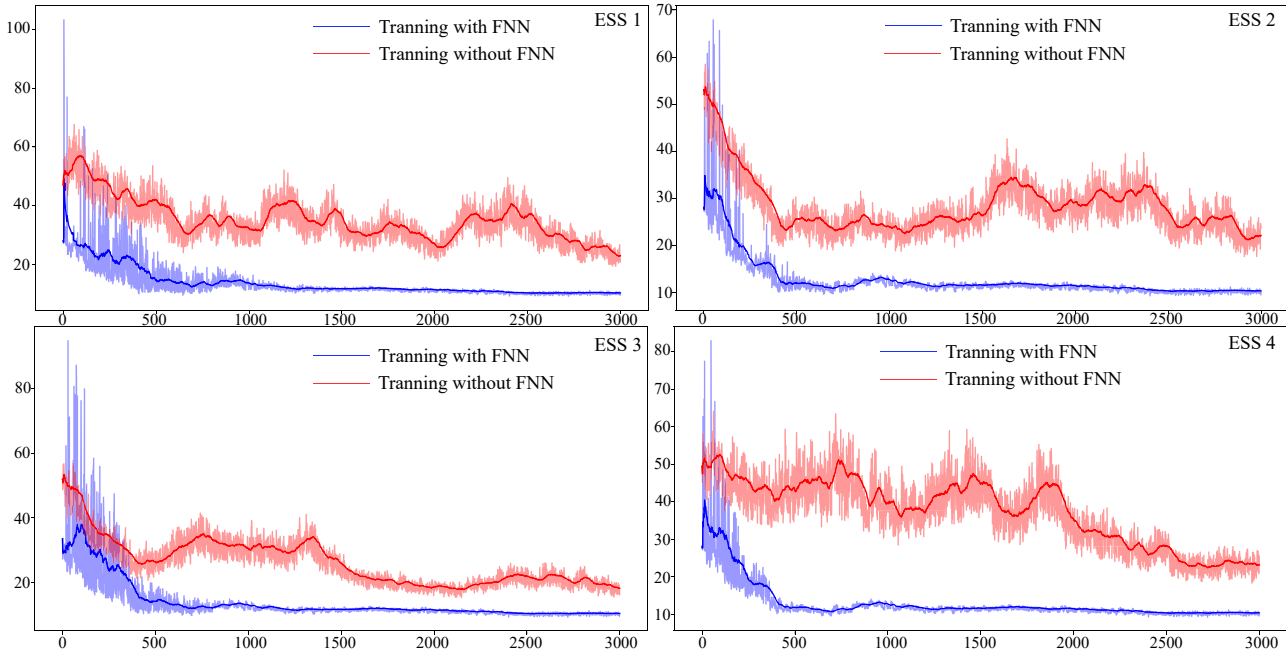
- **Gradient Leveling of the Loss Function**: Apply gradient leveling near the interaction sequence in the loss function to minimize the sensitivity.
  梯度正则化

- **New Neural Network for Enhanced Convergence**: handle uncertainty and noise in input variables by incorporating new neural network, enhancing the convergence ability.
  新型深度结构以提升收敛性

- **Decision-Making Strategy**: FDI attacks occur less frequently, a hybrid decision-making strategy is used:

$$\boldsymbol{a_t} = \begin{cases} \mu_{\Pi}(\boldsymbol{s}_t), & \mathcal{R}\left(\boldsymbol{s}_t, \mu_{\Pi}(\boldsymbol{s}_t)\right) = 1 \\ \mu_{\Pi^G}(\boldsymbol{s}_t), & \mathcal{R}\left(\boldsymbol{s}_t, \mu_{\Pi}(\boldsymbol{s}_t)\right) = 0 \end{cases}$$

混合策略提高精度



An example using the FNN structure .

Conclusions & Future Work

Conclusions

➢ Emerging Risks under DRL-Driven ADNs

Open-edge sensors + local gradient sensitivity, create new risks.

➢ Strong Impact of Adversarial Attacks

Both continuous-domain and discrete-domain attacks can cause severe operational risks with only small perturbations.

➢ Layered Defense Strategies Enhance Robustness

Adversarial training, curriculum learning training (CAT), and network structure enhancements provides substantial improvements in dispatch resilience under adversarial environments.

Future Work

➢ Extension to Integrated Energy Systems (IESs)

Expand the adversarial robustness framework to multi-energy systems, including electricity, heating, and cooling networks.

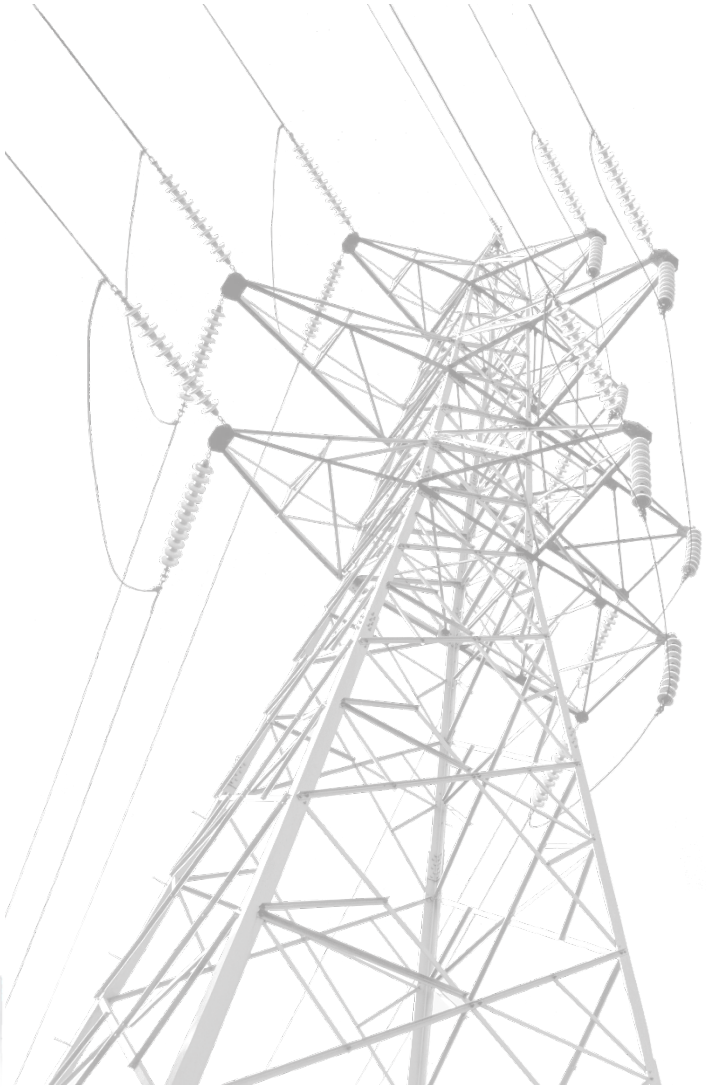
➢ DRL-Driven Real-Time Vulnerability Assessment

Develop real-time methods to dynamically identify and quantify system vulnerabilities during DRL-based dispatch operations.

➢ Advanced Défense Mechanism Development

Design adaptive defense-switching strategies and explore co-optimization of control and communication infrastructures for greater system resilience.

Thanks



Dr. SU YI