Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Smart Grid and Energy Internet—Article

利用分布式电源针对智能电网中数据驱动算法的虚假数据注入攻击方法

刘增稷[a]，刘梦歌[b,*]，王琦[c]，汤奕[c]

[a] Nanjing University of Posts and Telecommunications, Nanjing 210000, China
[b] Nanjing University, Nanjing 210023, China
[c] Southeast University, Nanjing 210000, China

ARTICLE INFO

Article history:
Received 24 March 2024
Revised 7 August 2024
Accepted 14 November 2024
Available online 6 December 2024

关键词
网络安全
数据驱动
网络攻击
生成对抗网络

摘要

随着用户侧分布式电源数量的增加，智能电网正变得越来越庞大和复杂。在数据驱动控制方法广泛应用的背景下，这些变化给智能电网带来了新的安全挑战。本文提出了一种新型黑盒虚假数据注入攻击（false data injection attack, FDIA）方法，本方法利用智能电网中分布式电源的测量模块，能够有效绕过常规安全措施。与传统方法聚焦于通信网络内的数据篡改不同，本方法直接在测量点注入虚假数据，使用生成对抗网络（generative adversarial network, GAN）生成隐蔽的攻击向量。本方法无需目标系统的详细知识，因而在实际攻击场景中具有较强的可行性。实验结果验证了这一攻击方法对电力系统稳定性的影响，凸显了数据驱动算法在智能电网中引入的重大网络安全风险。

© 2024 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. 引言

智能电网用户侧大量接入电池储能系统（battery energy storage system, BESS）和光伏（photovoltaic, PV）等分布式电源，为电网的清洁高效运行提供了重要支撑[1–2]。可控设备数量的快速增加使电力系统管理日趋复杂，传统的基于模型的分析和控制方法面临严峻挑战，系统与数据规模的持续扩张使得最优解的推导愈发困难[3–4]。数据驱动算法凭借其强大的数据处理能力，成为应对这一挑战的重要技术手段[5–9]。

智能电网向数字化、信息化和智能化方向演进的过程中，数据驱动算法面临着重大安全挑战，基于人工智能（artificial intelligence, AI）的算法尤为突出[10]。这些算法在离线训练和在线部署阶段对数据质量的要求极为严格[11–12]。攻击者可在训练期间通过操纵训练数据嵌入后门，或在部署阶段利用精心设计的攻击向量破坏算法功能[13–14]。数据驱动算法的应用为电网带来了新的网络安全风险[15–16]。用户侧分布式电源普遍依赖物联网（internet of things, IoT）设备，这类设备的网络防御能力相对薄弱，极易成为攻击目标[17]。相较于直接攻击发电厂或核心基础设施，针对用户侧设备的网络攻击虽然表面影响较小，但可能对系统运行产生深远影响[18–19]。智能电网环境下数据驱动方法的网络安全问题亟待深入研究。

虚假数据注入攻击（false data injection attack, FDIA）通过劫持或篡改数据损害系统运行，已成为智能电网面临

* Corresponding author.
E-mail address: 15366075521@163.com (M. Liu).

2095-8099/© 2024 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
英文原文：Engineering 2025, 51(8): 62–74
引用本文：Zengji Liu, Mengge Liu, Qi Wang, Yi Tang. False Data Injection Attacks on Data-Driven Algorithms in Smart Grids Utilizing Distributed Power Supplies. *Engineering*, https://doi.org/10.1016/j.eng.2024.11.025

的主要威胁之一[20–22]。近期研究集中探讨了利用最少资源注入攻击向量并欺骗状态估计过程的策略[23]。文献[4,24–25]讨论了易受入侵节点的选择方法。部分研究提出了仅需局部信息即可破坏直流（direct current, DC）状态估计的攻击方法[26]，文献[27]则提出了有效针对交流（alternating current, AC）状态估计的攻击技术。文献[28]描述的框架利用FDIA通过改变测量数据欺骗控制中心，构成了所谓的网络拓扑攻击。文献[29]针对基于共识的能量管理算法开展了攻击研究，通过操纵系统工作点实现累积性破坏效果。

文献[30]探索了在智能电网中部署生成对抗网络（generative adversarial network, GAN）执行数据驱动FDIA的方法。该方法利用GAN的生成能力复制正常运行数据的统计分布，产生难以检测的异常数据，但其研究重点仍主要停留在网络通信层的数据操纵。文献[31]在此基础上开发了深度时间卷积GAN（deep temporal convolutional GAN, DTCGAN）方法，强调时间关系在生成时间序列数据以模拟电网测量中的重要性，但在很大程度上仍局限于传统网络攻击方法。文献[32]引入的GAN框架测试了GAN在模拟和破坏智能电网稳定性预测方面的功效，展示了GAN误导稳定性预测系统的潜力。

电力系统暂态稳定预测（transient stability prediction, TSP）策略常成为破坏电网运行的攻击目标。文献[33]审查了FDIA对基于机器学习（machine learning, ML）的电力系统动态安全评估模型的影响，引入了鲁棒性验证方法来量化模型对FDIA的弹性。文献[34]探讨了基于ML的短期电压稳定性评估方法对FDIA的敏感性，提出了通过对抗训练加强ML模型防御能力的缓解策略。

现有研究主要针对状态估计和其他模型驱动算法开展攻击分析，针对智能电网内数据驱动算法的攻击研究则主要集中在电力市场领域，攻击目的在于获取经济利益或间接影响电网运行[35–36]。现有研究通常假设通信网络内传输的数据易被攻击者篡改，但虚假数据的注入过程往往未得到充分探讨。实际电力系统中，数据包通常通过光纤传输，关键数据采用加密保护，控制命令需要严格的身份验证。攻击者通过操纵通信网络内的数据包执行FDIA面临诸多技术障碍。针对电力系统中数据驱动算法的网络攻击，攻击者对潜在目标的信息访问往往受到限制，这一实际约束在许多研究分析中被忽视。

相比之下，分布式电源的测量数据通常以模拟信号或明文形式传输，安全防护相对薄弱。攻击者可通过在变压器或测量数据包中引入干扰损害分布式电源的控制，进而实现攻击向量向电力系统的注入。

在智能电网持续演进的背景下，数据驱动算法的安全性已成为亟待解决的关键问题之一。本文提出了一种基于黑盒范式的新型FDIA方法，攻击实施无需依赖目标算法的内部工作原理，仅通过对可观察输入和输出的操纵即可完成。现有FDIA方法多聚焦于对通信网络漏洞的利用，本研究则首次提出直接针对分布式电源测量模块的攻击途径。这一入口点在电力系统中安全性相对较低且长期未受重视。本方法符合实际网络安全场景的特征，即攻击者对专有或复杂系统细节的访问能力有限。与现有方法相比，本方法具备实时生成攻击向量的能力，虽需较长的离线训练时间，且仅能通过篡改局部区域内少数节点数据实施攻击。作为典型的黑盒攻击，本方法无需全面了解目标内部机制或分布式能源资源的详细规格，攻击者仅需访问目标的输入输出数据并破坏分布式电源的测量模块。本研究聚焦于针对智能电网数据驱动算法的攻击策略分析，详细阐述攻击方法，防御措施的探讨不在本文研究范围内。

本方法通过实时调整与物理参数集成生成和合法运行数据高度相似的攻击向量，对传统检测系统构成严峻挑战。文献[37]提出的A3D模型采用基于注意力机制的自动编码器，通过重建正常运行条件并与当前状态比较来识别异常。该模型对采用可预测模式修改数据的典型FDIA策略较为有效，但本文所提方法与物理参数的深度集成能够更准确地模仿合法运行数据，可规避此类基于重建的异常检测机制。文献[38]采用深度自动编码器进行窃电检测，其原理与FDIA检测类似，但面对与物理系统行为高度一致的篡改数据时，检测有效性可能显著降低。文献[39]使用的图神经网络（graph neural network, GNN）利用电网拓扑结构检测异常，在基于网络特征识别异常模式方面表现出色。本文所述方法通过分布式电源注入物理上合理的数据点，可削弱此类基于网络拓扑的检测机制的功效。文献[40]提出的扩展GNN方法可同时实现异常检测与定位，但面对反映真实运行变化特征的篡改数据时，检测和定位的准确性都可能受到影响。

本研究的主要贡献体现在三个方面：（1）基于分布式电源的常规建模，提出了控制器和滤波器参数的估计方法，使得黑盒攻击策略无需获取目标系统内部机制的详细知识，仅需分析输入输出数据即可实施攻击，显著简化了攻击执行的技术难度；（2）利用分布式电源测量模块，提出了一种新型虚假数据注入技术，能够有效误导状态估计过程，且在电力系统中具有良好的实施可行性；（3）通过设计攻击向量的分段生成过程，有效降低了计算开销并提高了执行效率，同时缓解了参数估计误差对攻击成功率的

不利影响。

本文后续内容如下：第2节介绍以GAN为基础框架的攻击向量生成方法；第3节构建基于分布式能源资源的FDIA模型；第4节概述完整的攻击实施过程；第5节通过实证研究评估所提攻击技术的有效性和性能表现；第6节总结研究成果并展望未来研究方向。

2. 基于GAN的攻击向量生成方法

本文阐述的攻击方法聚焦于电力系统框架内采用的数据驱动方法，本方法在训练完成后，可在超平面上构建决策边界，从而实现样本分类。如果使用此攻击方法生成精心设计的干扰并将其注入电力系统，则可能将某些样本推过边界，导致策略误判。

2.1. 理论假设

本文专注于利用分布式电源将攻击向量注入电力系统。因此，为了简化获取目标系统内控制策略的过程，做如下假设。

（1）攻击目标的复杂架构和控制机制是模糊的，但可以实现对目标输入和输出接口的监视，因为对于攻击者而言，拦截数据比操纵数据要容易得多。

（2）攻击目标的输入由电力系统的测量数据组成，或与测量数据线性相关。

（3）某些用户侧分布式电源可能被攻击者破坏，允许泄露控制器的结构。因此，攻击者可以访问和操纵由电压互感器（potential transformer, PT）和电流互感器（current transformer, CT）产生的测量数据。

（4）本文在攻击期间不考虑天气、温度等环境变量对分布式电源输出的影响。

2.2. 问题描述

生成攻击向量可以作为优化问题处理：获取最小扰动r，使其与原始数据叠加时，能够使攻击目标的输出偏离其标准度量。优化目标是以尽可能小的数据改动实现最大程度的破坏效果，同时规避常规安全检测机制。

$$\min\ Z=L(X,X+r)-\eta\log\prod_{j_a\in S_{i_a}}p_{j_a} \tag{1}$$

$$\text{s.t.}\ -T_B\le r\le T_B \tag{2}$$

$$\left|F(X+r)-F(X)\right|\ge\varepsilon \tag{3}$$

$$S_{i_a}\subseteq \text{SP} \tag{4}$$

$$v_{j_a\in S_{i_a}}=1,\ v_{j_a\notin S_{i_a}}=0 \tag{5}$$

$$\sum_{v_{j_a\in S_{i_a}}}v_{j_a}<\text{MAXA} \tag{6}$$

式中，Z是优化目标；L表示相似性评估指标，用于衡量扰动数据$X+r$与原始系统数据X的接近程度，确保扰动不易被检测到，它使用基于范数的度量来计算，该度量量化原始和更改数据向量之间的欧几里得距离，并根据系统对扰动的容忍度进行调整；$\eta\log\prod_{j_a\in S_{i_a}}p_{j_a}$用于平衡扰动的影响和成功攻击的可能性，其中$\eta$是常数，$p_{j_a}$是成功破坏可攻击节点集$S_{i_a}$中每个节点的概率（$i_a$是集合的编号，$j_a$是节点的编号），该组件确保攻击聚焦于系统中最脆弱的点，从而最大化效率；变量T描绘了干扰幅度的上限；$F(\cdot)$表征了支配攻击目标输入和输出之间关系的转换函数；符号ε表示攻击结果与既定规范度量的最小必要偏差；SP是易受攻击的分布式电源的集合；v_{j_a}表示节点j_a是否被攻击，其中1表示被攻击者控制，0表示正常运行；MAXA是攻击者可以控制的分布式电源的最大数量。

2.3. 攻击模型

考虑到攻击资源有限，攻击者通常只能渗透和控制节点的子集，并且这些节点的保护资源和策略各不相同，导致攻击者对不同节点的成功攻击概率不同。因此，为了实现预期攻击目标，攻击者必须考虑包括攻击资源、成功概率和攻击向量的干扰幅度在内的多种因素组合，以选择最佳攻击方案。在智能电网中选择漏洞进行攻击的过程可以通过以下优化问题来描述：如何在满足攻击资源、攻击向量生成和攻击效果约束的同时，将攻击拦截概率P最小化。

$$\min P=1-\prod_{k_a=1}^{N_a}v_{k_a}p_{k_a} \tag{7}$$

$$\text{s.t.}\ \text{ib}\subseteq\text{SA} \tag{8}$$

$$v_{k_a\in\text{ib}}=1,\ v_{k_a\notin\text{ib}}=0 \tag{9}$$

$$\sum_{k_a=1}^{N_a}v_{k_a}<\text{MAXA} \tag{10}$$

$$r=z(x,v) \tag{11}$$

$$\left|f\left(y(t),F\left(x(t)+v\times r\right)\right)-f\left(y(t),F\left(x(t)\right)\right)\right|\ge\gamma \tag{12}$$

式中，N_a是节点总数；k_a是节点编号，是1~N_a的整数；p_{k_a}是攻击节点k_a的成功概率；SA表示易受攻击的节点集；ib表示攻击者从SA中选择的易受攻击节点的子集。$z(\cdot)$是生成攻击向量的方法；v_{k_a}是节点k_a的攻击状态；x是原始测量数据；$x(t)$表示时间t的系统测量值；$y(t)$表示时间t的系统状态；$f(\cdot)$描述数据驱动算法影响下的系统状态；γ是攻击者希望对系统状态实现的最小影响。通过求解此优化问题，可以获得满足攻击目标的最优攻击策略。

每个节点的入侵过程可以使用Petri网建模，从而计算每个节点的攻击成功概率。基本方法涉及独立建模攻击和防御过程，分解攻击步骤，并计算每个步骤的发生概率。基于系统拓扑和数据传输路径构建攻击过程的综合模型，最终计算得到攻击成功概率。使用Petri网计算攻击成功概率的方法不是本研究的重点或本文的新贡献，因此不详细阐述，本文假设攻击者已经根据每个节点的防御资源计算了攻击成功概率。

在我们之前的研究[41]中，采用GAN来解决上述问题。在所描述的方案中，生成器G负责生成扰动数据，而判别器D负责评估其可检测性，通过交替迭代训练最终得到能够输出符合约束条件的攻击向量生成器。

考虑到攻击者只能成功破坏和控制分布式电源的子集，并且攻击的计算时间必须显著低于测量时间窗口，G的架构被特意设计为图1所示。该结构包括四个组件：节点选择器、编码器、解码器和滤波器。节点选择器的任务是选择最可行的攻击策略，编码器提取系统特征和每个节点数据之间的相互关系，解码器负责创建符合幅度边界条件的干扰数据源，滤波器产生最终扰动r。

在训练阶段结束后，可以将电网的常规数据实时馈送到生成器中，促进攻击向量$\bar{X}$的生成。在传统的FDIA中，$\bar{X}$通常通过通信网络引入电力系统。然而，在本文所述方法中，$\bar{X}$通过分布式电源注入，详见以下小节。

3. FDIA模型的构建

上述方法生成的攻击向量能够有效误导攻击目标。相较于从用户侧IoT设备注入，通过具备完善保护措施的通信网络注入攻击向量难度极大。因此，本文提出了一种基于分布式电源的虚假数据注入方法。为实现攻击向量的精确注入，需要估计分布式电源的参数并计算篡改后的测量序列。

3.1. 变流器模型

图2展示了分布式电源的典型结构，包括四个关键组件：直流电源、DC/DC变换器、DC/AC变换器和电感-电容-电感（inductor-capacitor-inductor, LCL）滤波器。直流电源可由电池或光伏板等组成，DC/DC控制器中采用下垂控制算法以均衡每个电源的输出。DC/AC变换器与LCL滤波器配合，主要完成直流到交流的电能转换。

LCL滤波器的特征由以下方程描述：

$$\begin{bmatrix} i'_{d,q} \\ u'_{d,q} \\ i'_{gd,q} \end{bmatrix} = \boldsymbol{A}_{\mathrm{LCL}} \times \begin{bmatrix} i_{d,q} \\ u_{d,q} \\ i_{gd,q} \end{bmatrix} + \boldsymbol{B}_{\mathrm{LCL}} \times \begin{bmatrix} u^*_{d,q} \\ u_{gd,q} \end{bmatrix} \tag{13}$$

$$\boldsymbol{A}_{\mathrm{LCL}} = \begin{bmatrix} -\frac{R_{\mathrm{f}}}{L_{\mathrm{f}}} & \omega & -\frac{1}{L_{\mathrm{f}}} & 0 & 0 & 0 \\ -\omega & -\frac{R_{\mathrm{f}}}{L_{\mathrm{f}}} & 0 & -\frac{1}{L_{\mathrm{f}}} & 0 & 0 \\ \frac{1}{C_{\mathrm{f}}} & 0 & 0 & \omega & -\frac{1}{C_{\mathrm{f}}} & 0 \\ 0 & \frac{1}{C_{\mathrm{f}}} & -\omega & 0 & 0 & -\frac{1}{C_{\mathrm{f}}} \\ 0 & 0 & \frac{1}{L_{\mathrm{g}}} & 0 & -\frac{R_{\mathrm{g}}}{L_{\mathrm{g}}} & \omega \\ 0 & 0 & 0 & \frac{1}{L_{\mathrm{g}}} & -\omega & -\frac{R_{\mathrm{g}}}{L_{\mathrm{g}}} \end{bmatrix} \tag{14}$$

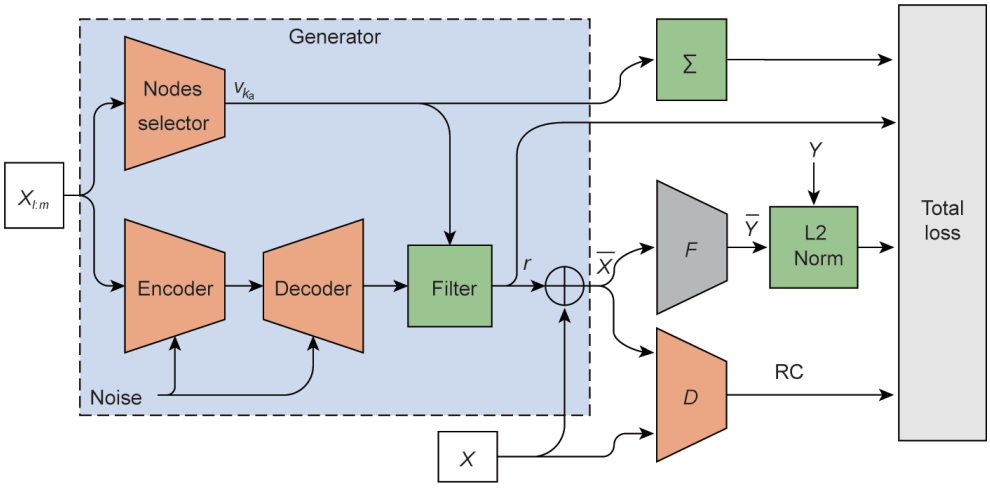

图1. 攻击向量生成器的训练过程。X：原始系统数据；l:m：从l到m的窗口；$X_{l:m}$：时间窗口l:m中的测量数据；∑：计算所有选定节点造成的总损失；F：攻击目标；$\bar{X}$：攻击向量；$\bar{Y}$：输入攻击向量时攻击目标的响应；Y：输入正常数据时攻击目标的响应；L2 Norm：Y和$\bar{Y}$之间的距离；D：攻击向量的判别器；RC：攻击向量$\bar{X}$的判别结果。

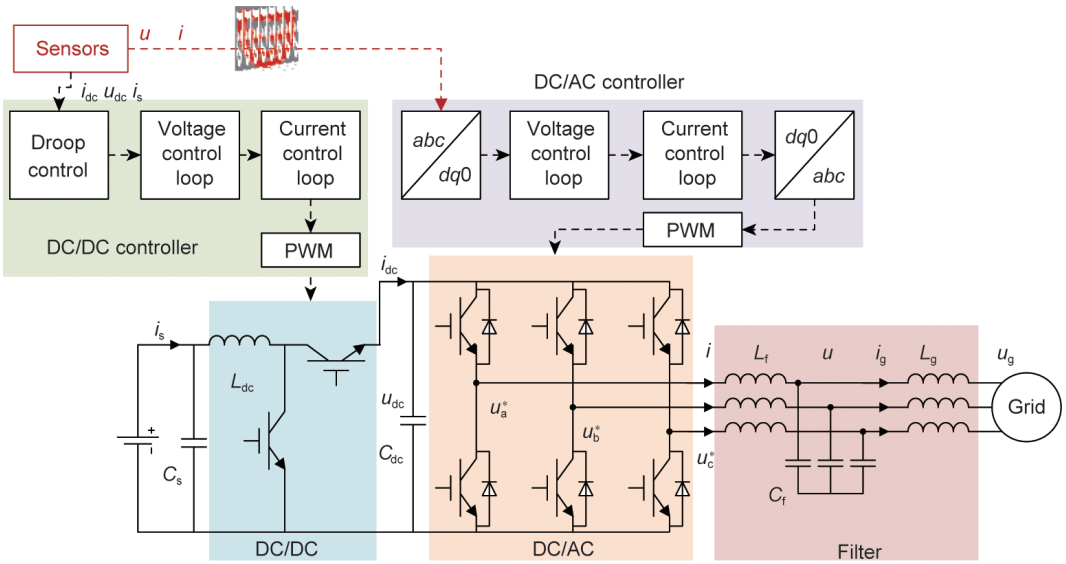

图2. 分布式电源的典型结构。i_s和C_s：分别为直流电源的电流和电容；L_{dc}：直流电源的电感；L_f和L_g：滤波器的电感；C_f：滤波器的电容；C_{dc}、u_{dc}和i_{dc}：分别为DC/AC变换器直流侧的电容、电压和电流；u_a^*、u_b^*和u_c^*：分别为DC/AC变换器a、b和c相的电压；u：滤波器中电容C_f的电压；i：电感L_f的电流；u_g：电网侧的电压；i_g：流入电网的电流；abc：abc坐标系中的值；$dq0$：$dq0$坐标系中的值；PWM：脉宽调制。

$$\boldsymbol{B}_f=\begin{bmatrix}\frac{1}{L_f} & 0\\ 0 & \frac{1}{L_f}\end{bmatrix} \tag{15}$$

$$\boldsymbol{B}_g=\begin{bmatrix}-\frac{1}{L_g} & 0\\ 0 & -\frac{1}{L_g}\end{bmatrix} \tag{16}$$

$$\boldsymbol{B}_{LCL}=\begin{bmatrix}\boldsymbol{B}_f & 0\\ 0 & 0\\ 0 & \boldsymbol{B}_g\end{bmatrix} \tag{17}$$

式中，$u_{d,q}^*$表示DC/AC变换器的输出电压；$i_{d,q}$表示逆变侧的电流；$u_{gd,q}$和$i_{gd,q}$分别表示电网侧的电压和电流；$u_{d,q}$表示滤波电容C_f两端的电压。R_f和L_f是逆变侧LCL滤波器的电阻和电感；R_g和L_g是电网侧滤波器的电阻和电感。ω表示电力系统的频率，0表示维度为2×2的零矩阵。$i'_{d,q}$、$u'_{d,q}$和$i'_{gd,q}$分别是$i_{d,q}$、$u_{d,q}$和$i_{gd,q}$的导数，$\boldsymbol{A}_{LCL}$和$\boldsymbol{B}_{LCL}$是滤波器的参数矩阵。$\boldsymbol{B}_f$表示滤波器侧的参数，$\boldsymbol{B}_g$表示电网侧的参数。

3.2. 控制器模型

在传感装置内，来自主电路的电压和电流分别通过PT和CT传送到副电路。然后，测量模块以固定频率采集数据点，通过快速傅里叶变换（fast fourier transform, FFT）计算电压和电流：

$$\begin{cases} i_R=\frac{2}{N}\sum_{l=1}^{N} i(l)\cos l\frac{2\pi}{N}\\ i_X=\frac{2}{N}\sum_{l=1}^{N} i(l)\sin l\frac{2\pi}{N}\\ u_R=\frac{2}{N}\sum_{l=1}^{N} u(l)\cos l\frac{2\pi}{N}\\ u_X=\frac{2}{N}\sum_{l=1}^{N} u(l)\sin l\frac{2\pi}{N}\end{cases} \tag{18}$$

$$\begin{cases} I=\sqrt{\frac{{i_R}^2+{i_X}^2}{2}}\\ \theta_I=\tan^{-1}\frac{i_X}{i_R}\\ U=\sqrt{\frac{{u_R}^2+{u_X}^2}{2}}\\ \theta_U=\tan^{-1}\frac{u_X}{u_R}\end{cases} \tag{19}$$

式中，N表示样本总数；$i(l)$和$u(l)$分别表示第l个采样点的电流和电压；i_R、i_X、u_R和u_X分别是通过FFT计算的电流实部、电流虚部、电压实部和电压虚部值。因此，可以根据式（18）和式（19）计算出三相电压$u_{a,b,c}$和三相电流$i_{a,b,c}$，其中I是电流的有效值，θ_I是电流的角度，U是电压的有效值，θ_U是电压的角度。

第2节讨论的基于GAN的攻击策略只需要向原始数据注入微小干扰即可使目标失效，干扰通常表现为短暂时间内的一系列电压脉冲。因此，在攻击期间，可以通过操纵DC/AC变换器将攻击向量注入电力系统。图3为DC/AC变换器控制器的标准结构，采用电压控制环和电流控制环来调节变换器的输出。

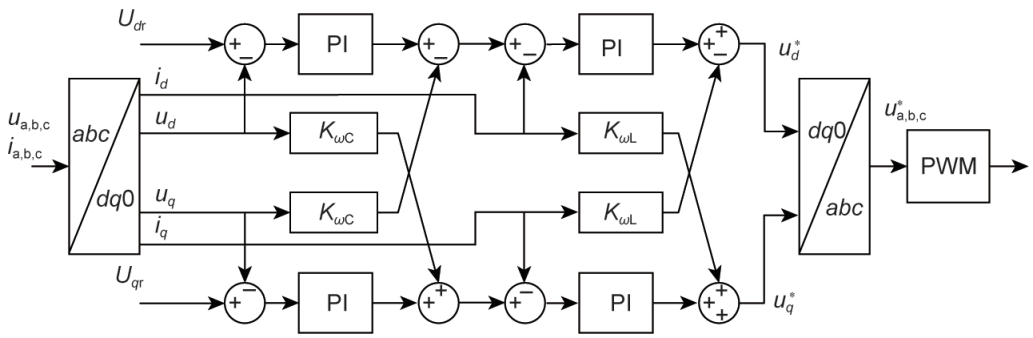

图3. DC/AC控制器的典型结构。PI：比例-积分；$u_{a,b,c}$和$i_{a,b,c}$：三相中的电压和电流；u_d和u_q：沿d-q轴的电压分量；i_d和i_q：沿d-q轴的电流分量；U_{dr}和U_{qr}：沿d-q轴的参考电压；$K_{\omega C}$和$K_{\omega L}$：与DC/AC控制器相关的参数；u_d^*和u_q^*：沿d-q轴由控制器输出的电压，可通过反Park变换转换为$u_{a,b,c}^*$。

沿d-q轴的电压u_d和u_q、电流i_d和i_q可分别通过Park变换从$u_{a,b,c}$和$i_{a,b,c}$计算得到。U_{dr}和U_{qr}表示参考电压，$K_{\omega C}$和$K_{\omega L}$是与DC/AC控制器相关的参数。比例-积分（proportional-integral, PI）控制器通常在电压和电流控制环中使用。u_d^*和u_q^*是沿d-q轴由控制器输出的电压，可通过反Park变换转换为$u_{a,b,c}^*$。上述参数可以离散化如下：存在$t=k\tau$，其中τ表示固定步长，k表示步数。

$$O_{\mathrm{PI}}(k)=K_{\mathrm{P}}\left[e(k)+\frac{1}{T_{\mathrm{I}}}\sum_{j=0}^{k}Te(j)\right]=K_{\mathrm{P}}e(k)+K_{\mathrm{I}}\sum_{j=0}^{k}e(j) \tag{20}$$

$$\Delta O_{\mathrm{PI}}(k)=K_{\mathrm{P}}\left[e(k)-e(k-1)\right]+K_{\mathrm{I}}e(k) \tag{21}$$

式中，e和O_{PI}分别对应PI控制器的输入和输出；K_{P}和K_{I}分别是PI控制器中比例和积分分量的参数；T_{I}是积分时间常数；T是每个步长内的时间。因此，DC/AC变换器的输出电压$u_{d,q}^*$与测量数据$u_{d,q}$和$i_{d,q}$之间的函数关系如下：

$$\begin{bmatrix}\Delta u_d^*(k)\\ \Delta u_q^*(k)\end{bmatrix}=\begin{bmatrix}\Delta u_d^*(k-1)\\ \Delta u_q^*(k-1)\end{bmatrix}+K_{\mathrm{Ii}}K_{\mathrm{Iu}}\begin{bmatrix}U_{dr}\\ U_{qr}\end{bmatrix}+\begin{bmatrix}K_d\\ K_q\end{bmatrix}M(k) \tag{22}$$

$$\begin{bmatrix}u_d^*(k)\\ u_q^*(k)\end{bmatrix}=\begin{bmatrix}u_d^*(k-1)\\ u_q^*(k-1)\end{bmatrix}+\begin{bmatrix}\Delta u_d^*(k)\\ \Delta u_q^*(k)\end{bmatrix} \tag{23}$$

初始值配置为系统的稳态值，u_{ds}^*和u_{qs}^*表示逆变侧的稳态电压。

$$\begin{bmatrix}u_d^*(0)\\ u_q^*(0)\end{bmatrix}=\begin{bmatrix}u_{ds}^*\\ u_{qs}^*\end{bmatrix},\begin{bmatrix}\Delta u_d^*(0)\\ \Delta u_q^*(0)\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix} \tag{24}$$

$M(k)$表示从第k步到第$k-2$步的测量数据，包括沿d轴和q轴的电压和电流，其计算方法如下：

$$M(k)=\left[M_{\mathrm{u}d}(k),M_{\mathrm{u}q}(k),M_{\mathrm{i}d}(k),M_{\mathrm{i}q}(k)\right]^{\mathrm{T}} \tag{25}$$

$$M_{\mathrm{u}d}(k)=\left[u_d(k),u_d(k-1),u_d(k-2)\right] \tag{26}$$

$$M_{\mathrm{u}q}(k)=\left[u_q(k),u_q(k-1),u_q(k-2)\right] \tag{27}$$

$$M_{\mathrm{i}d}(k)=\left[i_d(k),i_d(k-1),i_d(k-2)\right] \tag{28}$$

$$M_{\mathrm{i}q}(k)=\left[i_q(k),i_q(k-1),i_q(k-2)\right] \tag{29}$$

式中，$M_{\mathrm{u}d}(k)$和$M_{\mathrm{u}q}(k)$是d轴和q轴测量电压值序列；$M_{\mathrm{i}d}(k)$和$M_{\mathrm{i}q}(k)$是d轴和q轴测量电流值序列。

K_d和K_q是调节d轴和q轴的控制参数，如下所示：

$$K_d=\begin{bmatrix}1-K_{\mathrm{Pi}}K_{\mathrm{Pu}}-K_{\mathrm{Pi}}K_{\mathrm{Iu}}-K_{\mathrm{Ii}}K_{\mathrm{Pu}}-K_{\mathrm{Ii}}K_{\mathrm{Iu}}\\ -2+2K_{\mathrm{Pi}}K_{\mathrm{Pu}}+K_{\mathrm{Pi}}K_{\mathrm{Iu}}+K_{\mathrm{Ii}}K_{\mathrm{Pu}}\\ 1-K_{\mathrm{Pi}}K_{\mathrm{Pu}}\\ -K_{\omega C}K_{\mathrm{Pi}}-K_{\omega C}K_{\mathrm{Ii}}\\ 2K_{\omega C}K_{\mathrm{Pi}}+K_{\omega C}K_{\mathrm{Ii}}\\ -K_{\omega C}K_{\mathrm{Pi}}\\ -K_{\mathrm{Pi}}-K_{\mathrm{Ii}}\\ 2K_{\mathrm{Pi}}+K_{\mathrm{Ii}}\\ -K_{\mathrm{Pi}}\\ -K_{\omega L}\\ 2K_{\omega L}\\ -K_{\omega L}\end{bmatrix}^{\mathrm{T}} \tag{30}$$

$$K_q=\begin{bmatrix}K_{\omega C}K_{\mathrm{Pi}}+K_{\omega C}K_{\mathrm{Ii}}\\ -2K_{\omega C}K_{\mathrm{Pi}}-K_{\omega C}K_{\mathrm{Ii}}\\ K_{\omega C}K_{\mathrm{Pi}}\\ 1-K_{\mathrm{Pi}}K_{\mathrm{Pu}}-K_{\mathrm{Pi}}K_{\mathrm{Iu}}-K_{\mathrm{Ii}}K_{\mathrm{Pu}}-K_{\mathrm{Ii}}K_{\mathrm{Iu}}\\ -2+2K_{\mathrm{Pi}}K_{\mathrm{Pu}}+K_{\mathrm{Pi}}K_{\mathrm{Iu}}+K_{\mathrm{Ii}}K_{\mathrm{Pu}}\\ 1-K_{\mathrm{Pi}}K_{\mathrm{Pu}}\\ K_{\omega L}\\ -2K_{\omega L}\\ K_{\omega L}\\ -K_{\mathrm{Pi}}-K_{\mathrm{Ii}}\\ 2K_{\mathrm{Pi}}+K_{\mathrm{Ii}}\\ -K_{\mathrm{Pi}}\end{bmatrix}^{T} \tag{31}$$

式中，K_{Pi}和K_{Ii}为电流控制环的PI控制器参数；K_{Pu}和K_{Iu}为控制电压控制环的PI控制器参数。脉宽调制（pulse width modulation, PWM）控制器根据$u_{d,q}^*$控制晶体管（insulated gate bipolar transistor, IGBT），以确保DC/AC变换器输出所需电压。

3.3. 参数估计

为了精确执行黑盒攻击注入攻击向量，攻击者必须估计分布式电源内滤波器和控制器的参数。在实际电力系统

中，测量数据通常包含误差ξ，其遵循正态分布$N\left(0,\sigma^2\right)$。因此，应使用加权最小二乘（weighted least squares, WLS）估计来估计参数β，以最小化$Y=X\beta$的残差，计算如下：

$$\arg\min_{\beta}\left\|\boldsymbol{W}\left(Y-X\beta\right)\right\|_2 \tag{32}$$

式中，$\boldsymbol{W}\in\mathbb{R}^{n\times n}$是权重矩阵；$n$是$Y$的长度；$Y$和$X$从测量数据中计算得到；$n_{\mathrm{w}}$表示在时间窗口内收集的数据点数。对于LCL滤波器，需要估计参数ω、L_{f}、R_{f}、C_{f}、L_{g}和R_{g}。因此，将式（13）至式（17）重新表示为$Y=X\beta$的形式，如下所示：

$$\left(i_q-i_{\mathrm{g}q}\right)u'_d-\left(i_d-i_{\mathrm{g}d}\right)u'_q=\left[u_d\left(i_d-i_{\mathrm{g}d}\right)+u_q\left(i_q-i_{\mathrm{g}q}\right)\right]\omega \tag{33}$$

$$u_d u'_q+u_q u'_d=\left[u_d\left(i_d-i_{\mathrm{g}d}\right)+u_q\left(i_q-i_{\mathrm{g}q}\right)\right]\frac{1}{C_{\mathrm{f}}} \tag{34}$$

$$i_q i'_d-i_d i'_q-\left(i_d{}^2+i_q{}^2\right)\omega=\left[\left(u_d^*-u_d\right)i_q-\left(u_q^*-u_q\right)i_d\right]\frac{1}{L_{\mathrm{f}}} \tag{35}$$

$$\begin{aligned}&\left(u_q^*-u_q\right)i'_d-\left(u_d^*-u_d\right)i'_q-\left[\left(u_d^*-u_d\right)i_d+\left(u_q^*-u_q\right)i_q\right]\omega=\\&\left[\left(u_d^*-u_d\right)i_q-\left(u_q^*-u_q\right)i_d\right]\frac{R_{\mathrm{f}}}{L_{\mathrm{f}}}\end{aligned} \tag{36}$$

$$\begin{aligned}&i_{\mathrm{g}q}i'_{\mathrm{g}d}-i_{\mathrm{g}d}i'_{\mathrm{g}q}-\left(i_{\mathrm{g}d}{}^2+i_{\mathrm{g}q}{}^2\right)\omega=\\&\left[\left(u_d-u_{\mathrm{g}d}\right)i_{\mathrm{g}q}-\left(u_q-u_{\mathrm{g}q}\right)i_{\mathrm{g}d}\right]\frac{1}{L_{\mathrm{g}}}\end{aligned} \tag{37}$$

$$\begin{aligned}&\left(u_q-u_{\mathrm{g}q}\right)i'_{\mathrm{g}d}-\left(u_d-u_{\mathrm{g}d}\right)i'_{\mathrm{g}q}-\\&\left[\left(u_d-u_{\mathrm{g}d}\right)i_{\mathrm{g}d}+\left(u_q-u_{\mathrm{g}q}\right)i_{\mathrm{g}q}\right]\omega=\\&\left[\left(u_d-u_{\mathrm{g}d}\right)i_{\mathrm{g}q}-\left(u_q-u_{\mathrm{g}q}\right)i_{\mathrm{g}d}\right]\frac{R_{\mathrm{g}}}{L_{\mathrm{g}}}\end{aligned} \tag{38}$$

式中，$i_{\mathrm{g}d}$和$i_{\mathrm{g}q}$是电网侧电流的d轴和q轴分量；u'_d、u'_q、$i'_{\mathrm{g}d}$和$i'_{\mathrm{g}q}$分别是u_d、u_q、$i_{\mathrm{g}d}$和$i_{\mathrm{g}q}$的导数。$u_{d,q}$和$i_{d,q}$可以从变换器侧的测量模块获得，$u^*_{d,q}$可从DC/AC控制器中获得，$u_{\mathrm{g}d,q}$和$i_{\mathrm{g}d,q}$可通过电网侧的测量模块获得。$i'_{d,q}$、$u'_{d,q}$和$i'_{\mathrm{g}d,q}$的离散化和近似可表示为：

$$\begin{bmatrix}i'_{d,q}(k)\\u'_{d,q}(k)\\i'_{\mathrm{g}d,q}(k)\end{bmatrix}=\frac{1}{\Delta t}\begin{bmatrix}i_{d,q}(k)-i_{d,q}(k-1)\\u_{d,q}(k)-u_{d,q}(k-1)\\i_{\mathrm{g}d,q}(k)-i_{\mathrm{g}d,q}(k-1)\end{bmatrix} \tag{39}$$

在DC/AC控制器内，需要估计参数K_d、K_q和C，其中C是控制器的常数，$C=K_{\mathrm{Ii}}K_{\mathrm{Iu}}[U_{dr}\quad U_{qr}]^{\mathrm{T}}$。因此，将式（22）重新表示为$Y=X\beta$的形式，如下所示：

$$\begin{bmatrix}\Delta u_d^*(k)-\Delta u_d^*(k-1)\\\Delta u_q^*(k)-\Delta u_q^*(k-1)\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}M(k)^{\mathrm{T}}&1\end{bmatrix}\times\begin{bmatrix}K_{d,q}{}^{\mathrm{T}}\\C^{\mathrm{T}}\end{bmatrix} \tag{40}$$

式中，$K_{d,q}=[K_d,K_q]$；$M(k)$来自变换器侧的测量模块；$\begin{bmatrix}\Delta u_d^*(k)\\\Delta u_q^*(k)\end{bmatrix}$使用式（23）计算。通过长期观察，攻击者能够精确估计LCL滤波器和DC/AC控制器的参数。

3.4. 攻击向量设计

第2节基于GAN生成的攻击向量可以显著破坏目标的正常功能，因此如果攻击者操纵分布式电源使其输出与攻击向量紧密对齐，则可以达到类似效果。攻击者可以修改测量模块采集的电流和电压数据序列，注入a_{i}和a_{u}，如式（18）所示：

$$\begin{cases}i_{\mathrm{RA}}=i_{\mathrm{R}}+\dfrac{2}{N}\sum_{l=1}^{N}a_{\mathrm{i}}(l)\cos l\dfrac{2\pi}{N}\\i_{\mathrm{XA}}=i_{\mathrm{X}}+\dfrac{2}{N}\sum_{l=1}^{N}a_{\mathrm{i}}(l)\sin l\dfrac{2\pi}{N}\\u_{\mathrm{RA}}=u_{\mathrm{R}}+\dfrac{2}{N}\sum_{l=1}^{N}a_{\mathrm{u}}(l)\cos l\dfrac{2\pi}{N}\\u_{\mathrm{XA}}=u_{\mathrm{X}}+\dfrac{2}{N}\sum_{l=1}^{N}a_{\mathrm{u}}(l)\sin l\dfrac{2\pi}{N}\end{cases} \tag{41}$$

式中，i_{RA}、i_{XA}、u_{RA}和u_{XA}分别是注入攻击向量后的电流实部、电流虚部、电压实部和电压虚部。因此，实施攻击可以重新定义为计算并修改测量数据问题，使设计的攻击向量与分布式电源输出结果之间的方差最小化，这一优化问题可以通过以下两个主要步骤解决。

3.4.1. 计算DC/AC变换器$u^*_{d,q\mathrm{A}}$的优化参考电压输出

鉴于测量数据和目标攻击向量均为离散数据，式（13）需要离散化如下：

$$\frac{1}{\Delta t}\begin{bmatrix}i_d(k)-i_d(k-1)\\i_q(k)-i_q(k-1)\end{bmatrix}=\boldsymbol{A}^{1:2}_{\mathrm{LCL}}\begin{bmatrix}i_d(k-1)\\i_q(k-1)\\u_d(k-1)\\u_q(k-1)\end{bmatrix}+\boldsymbol{B}_{\mathrm{f}}\begin{bmatrix}u_d^*(k)\\u_q^*(k)\end{bmatrix} \tag{42}$$

$$\frac{1}{\Delta t}\begin{bmatrix}u_d(k)-u_d(k-1)\\u_q(k)-u_q(k-1)\end{bmatrix}=\boldsymbol{A}^{3:4}_{\mathrm{LCL}}\begin{bmatrix}i_d(k)\\i_q(k)\\u_d(k-1)\\u_q(k-1)\\i_{\mathrm{g}d}(k-1)\\i_{\mathrm{g}q}(k-1)\end{bmatrix} \tag{43}$$

$$\begin{bmatrix}u_{\mathrm{g}d}(k)\\u_{\mathrm{g}q}(k)\end{bmatrix}=\begin{bmatrix}u_d(k-1)\\u_q(k-1)\end{bmatrix} \tag{44}$$

$$\frac{1}{\Delta t}\begin{bmatrix}i_{\mathrm{g}d}(k)-i_{\mathrm{g}d}(k-1)\\i_{\mathrm{g}q}(k)-i_{\mathrm{g}q}(k-1)\end{bmatrix}=\boldsymbol{A}^{5:6}_{\mathrm{LCL}}\begin{bmatrix}u_d(k)\\u_q(k)\\i_{\mathrm{g}d}(k-1)\\i_{\mathrm{g}q}(k-1)\end{bmatrix}+\boldsymbol{B}_{\mathrm{g}}\begin{bmatrix}u_{\mathrm{g}d}(k)\\u_{\mathrm{g}q}(k)\end{bmatrix} \tag{45}$$

式中，$\boldsymbol{A}_{\mathrm{LCL}}^{1:2}$ 表示由 $\boldsymbol{A}_{\mathrm{LCL}}$ 的第一行和第二行组成的矩阵，其初始值设置为系统的稳态值。基于上述方程，电网侧电压的变化可以由DC/AC变换器输出的改变确定。要计算 $u_{d,q\mathrm{A}}^{*}$，必须解决以下优化问题：

$$\min F_{\mathrm{LCL}}\left(u_{d,q\mathrm{A}}^{*}\right)=\left\|\mathrm{NC}\left(u_{\mathrm{g}d,q\mathrm{A}},u_{\mathrm{g}d,q\mathrm{T}}\right)-1\right\|_{2} \tag{46}$$

$$\text{s.t. } \mathrm{MIN}\le u_{d,q\mathrm{A}}^{*}\le \mathrm{MAX} \tag{47}$$

式中，$u_{\mathrm{g}d,q\mathrm{T}}$ 由 GAN 生成；$u_{\mathrm{g}d,q\mathrm{A}}$ 和 $u_{d,q\mathrm{A}}^{*}$ 由式（42）至式（45）计算得到；MIN和MAX定义DC/AC变换器输出的限制；$\mathrm{NC}\left(u_{\mathrm{g}d,q\mathrm{A}},u_{\mathrm{g}d,q\mathrm{T}}\right)$表示$u_{\mathrm{g}d,q\mathrm{A}}$和$u_{\mathrm{g}d,q\mathrm{T}}$之间的归一化相关性，计算如下所示。通过解决上述优化问题，可以得到DC/AC变换器的目标输出序列。

$$\mathrm{NC}=\frac{\frac{1}{N}\sum_{k=1}^{N}\left[u_{\mathrm{g}d\mathrm{A}}(k)\times u_{\mathrm{g}d\mathrm{T}}(k)+u_{\mathrm{g}q\mathrm{A}}(k)\times u_{\mathrm{g}q\mathrm{T}}(k)\right]}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[u_{\mathrm{g}d\mathrm{A}}(k)^{2}+u_{\mathrm{g}q\mathrm{A}}(k)^{2}\right]}\times\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[u_{\mathrm{g}d\mathrm{T}}(k)^{2}+u_{\mathrm{g}q\mathrm{T}}(k)^{2}\right]}} \tag{48}$$

3.4.2. 计算$u_{d,q\mathrm{A}}$和$i_{d,q\mathrm{A}}$

为了确定篡改的测量序列$u_{d,q\mathrm{A}}$和$i_{d,q\mathrm{A}}$，需要针对每个数据点最小化篡改值与正常测量序列之间的距离F_{ctl}，其中$x=\left[u_{d\mathrm{A}}(k)\quad u_{q\mathrm{A}}(k)\quad i_{d\mathrm{A}}(k)\quad i_{q\mathrm{A}}(k)\right]^{\mathrm{T}}$。

$$\min F_{\mathrm{ctl}}(x)=\sqrt{\frac{\sum\left(x_{1,2}-u_{d,q}(k-1)\right)^{2}}{W_{\mathrm{u}}}+\frac{\sum\left(x_{3,4}-i_{d,q}(k-1)\right)^{2}}{W_{\mathrm{i}}}} \tag{49}$$

$$\text{s.t. } K_{\min}\begin{bmatrix}u_{d}(k-1)\\u_{q}(k-1)\\i_{d}(k-1)\\i_{q}(k-1)\end{bmatrix}\le x\le K_{\max}\begin{bmatrix}u_{d}(k-1)\\u_{q}(k-1)\\i_{d}(k-1)\\i_{q}(k-1)\end{bmatrix} \tag{50}$$

$$K_{1}x+K_{2}\begin{bmatrix}u_{d}(k-1)\\u_{d}(k-2)\\u_{q}(k-1)\\u_{q}(k-2)\\i_{d}(k-1)\\i_{d}(k-2)\\i_{q}(k-1)\\i_{q}(k-2)\end{bmatrix}+C=\begin{bmatrix}\Delta u_{d}^{*}(k)-\Delta u_{d}^{*}(k-1)\\\Delta u_{q}^{*}(k)-\Delta u_{q}^{*}(k-1)\end{bmatrix} \tag{51}$$

$$\Delta u_{d,q}^{*}(k-1)=u_{d,q}^{*}(k-1)-u_{d,q}^{*}(k-2) \tag{52}$$

对于$u_{d,q}^{*}$，其初始值来自测量值，在每次迭代中，更新为式（46）优化结果$u_{d,q\mathrm{A}}^{*}$中的值。$u_{d,q}$和$i_{d,q}$的初始值同样来自测量值，在每次迭代中更新为前一次迭代的结果。W_{u}和W_{i}作为权重，用于平衡$u_{d,q\mathrm{A}}$和$i_{d,q\mathrm{A}}$对优化结果的影响。$K_{\min}$和$K_{\max}$约束篡改值与原始数据之间的差异。K_1和K_2是常系数矩阵，基于K_d和K_q计算如下：

$$K_{1}=\begin{bmatrix}K_{d1}&K_{d4}&K_{d7}&K_{d10}\\K_{q1}&K_{q4}&K_{q7}&K_{q10}\end{bmatrix} \tag{53}$$

$$K_{2}=\begin{bmatrix}K_{d2}&K_{d3}&K_{d5}&K_{d6}&K_{d8}&K_{d9}&K_{d11}&K_{d12}\\K_{q2}&K_{q3}&K_{q5}&K_{q6}&K_{q8}&K_{q9}&K_{q11}&K_{q12}\end{bmatrix} \tag{54}$$

式中，$K_{d\kappa}$和$K_{q\kappa}$分别为K_d和K_q的第κ个元素。

4. 攻击过程

电力系统的运行状态对FDIA的功效和可检测性具有显著影响，本文深入探究了系统运行状态对这类网络威胁的影响。在正常运行状态下，数据驱动控制系统具有固有冗余和鲁棒的错误检测机制，能够有效识别和缓解异常数据偏差，并将其归因于设备故障或传感器异常，从而降低攻击影响。然而，在系统紧急状态下，无论是人为故障还是自发事件，系统动态特性会发生实质性变化。此时，系统冗余机制可能被绕过或抑制，控制重心转向故障恢复和系统稳定。这种情况下，攻击向量逃避检测的可能性显著增加，其破坏潜力也大幅提升。在系统不稳定期间适时注入攻击向量可利用瞬态脆弱性，使故障后的系统更易受恶意数据影响。此外，此时注入的攻击向量可利用系统混乱状态，将异常归因于故障本身，从而增强攻击隐蔽性并最大化破坏效果，进一步导致控制策略做出错误决策，加剧系统不稳定性，甚至引发级联故障。

本文提出的攻击方法步骤如算法1所示，可概述如下。

步骤1：识别攻击目标。对保护不足的电源设置成功率p为1，对其余电源设置为0。选择目标F并评估电网内每个分布式电源的入侵难度，利用长期观测的样本训练GAN。

步骤2：参数估计。估计被攻击的分布式电源参数，通过式（32）至式（38）估计LCL滤波器的ω、L_{f}、R_{f}、C_{f}、L_{g}和R_{g}，并使用式（40）估计DC/AC控制器的K_d、K_q和C。

步骤3：监测系统异常事件。等待人为故障或自发紧急事件（如短路、停电等），在事件发生时使用GAN的生成器生成攻击向量$u_{\mathrm{g}d,q\mathrm{T}}$。

步骤4：计算最优DC/AC变换器输出。使用式（42）至式（48）确定DC/AC变换器输出的目标攻击向量$u_{d,q\mathrm{A}}^{*}$。

步骤5：生成篡改的测量序列。根据式（49）至式（54）计算得到篡改的测量序列$u_{d,q\mathrm{A}}$和$i_{d,q\mathrm{A}}$。

步骤6：注入干扰。使用式（41）计算干扰值a_{i}和a_{u}，并将其注入分布式电源的测量模块。

算法1 利用分布式电源的虚假数据注入

Input:

Information of the power system

Historical data of power supplies

Parameter constraints

Real-time measurement data of power system X

Model of the distributed power supply

Output:

Disturbance values a_i and a_u

Initialize:

Assign a success rate p for each distributed power supply

Initialize GAN with generator G and discriminator D

Initialize parameters of distributed power supply

Identify the attack target:

Define the optimization problem in Eqs. (7)–(12)

Solve the optimization problem using a heuristic algorithm

Find optimal attack target F

Assess the difficulty of intrusion for each distributed power supply

while GAN not converged do

Train D to differentiate between real operational data and generated data

Train G to create plausible attack data that fools D

end while

Parameter estimation:

Apply WLS to estimate parameters:

LCL filter parameters: ω, L_f, R_f, C_f, L_g, and R_g, using Eqs. (32)–(38)

Controller parameters: K_d, K_q, and C, using Eq. (40)

Validate parameters against benchmarks

Attack vector generation:

while power system operates normally

Await artificial faults or spontaneous emergencies

end while

Calculate the target attack vector $u^*_{d,q\mathrm{A}}$ using Eqs. (42)–(48)

Calculate the altered measurement sequence $u_{d,q\mathrm{A}}$ and $i_{d,q\mathrm{A}}$ using Eqs. (49)–(54)

Calculate the disturbance values a_i and a_u using Eq. (41)

Inject tampered data through the measurement modules of selected supplies

Monitor the system's response to the injection

攻击过程时间线如图4所示，攻击向量的生成和注入发生在测量时间窗口内。为了最小化误差并提高计算效率，本文采用分段过程生成攻击向量。攻击序列被分为多个段，每个段根据攻击阶段获得的最新测量数据进行重新校准和调整，从而显著减少初始参数估计误差在攻击过程中的累积效应。新数据的持续输入确保生成的攻击向量与系统的当前状态保持最佳对齐，从而减小由于参数估计不准确而被检出的可能性。

攻击过程分为α个阶段，如果t_w延长，则前一个阶段的攻击向量注入和下一阶段攻击向量的生成可以同时进行。在每个阶段，攻击者利用测量数据点1到h生成攻击向量，鉴于攻击方法需要计算时间，阶段1中的初始h个数据点保持不变，但所有后续数据点都可能被篡改。如果$\mathrm{num}-2h$过大，则表示α太小，攻击向量注入的精度会降低，可能降低攻击的成功率。相反，如果α过大以至于$(\mathrm{num}-2h)<0$，尽管攻击向量注入精度可能提高，但可篡改的数据点数量减少，攻击的成功率仍可能下降。因此，确定阶段数需要综合考虑测量时间窗口t_w和攻击向量注入所需的精度。

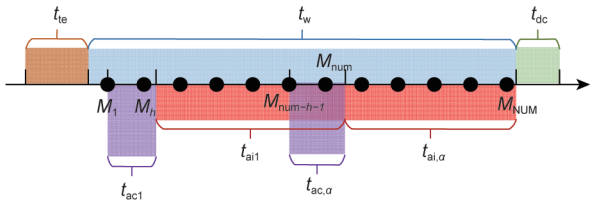

图4. 攻击时间线。t_{te}：目标F的触发事件持续时间；t_w：测量窗口；M：测量数据点，下标h为初始数据点，num为整个攻击序列中可供操作的数据点数，NUM为数据点总数；t_{ac1}和$t_{ac,\alpha}$：攻击者获取测量数据并创建攻击向量的时间段；t_{ai1}和$t_{ai,\alpha}$：攻击向量的注入时间；t_{dc}：攻击目标数据驱动算法在收集测量数据后的计算时间。

5. 算例分析

为了验证本研究提出的攻击策略的有效性，我们在新英格兰39节点系统场景中，针对基于深度卷积神经网络的TSP方法进行攻击。首先，概述攻击场景的构建；随后，展示DC/AC控制器和LCL滤波器参数的估计结果；最后，详细阐述并讨论攻击向量的生成及其后续影响。

5.1. 场景描述

在新英格兰39节点系统中，三个具有相同参数的分布式电源，连接到节点28、29和36。在每个节点部署相量测量单元（phasor measurement unit, PMU）以监测系统状态，如图5所示。负载功率在其基准值的80%~120%之间变化，三相短路故障可随机出现在任一线路长度的20%~80%处，故障的持续时间在0.10~0.35 s之间。为了模拟各种运行状态和故障场景，本研究采用蒙特卡洛法随机构建2000个样本数据，并将仿真确定的系统稳定性作为每个样本的标签。

在本场景中训练AlexNet模型，用于预测三相短路故障后电力系统的暂态稳定性。模型输入为故障后100 ms内的短时间窗口测量数据，输出为预测结果。训练完成后，模型的预测精度达到98%，表明其能够有效判定故障后电力系统的潜在不稳定性。假设攻击者已成功入侵连接到节点28、29和36的分布式电源，具备在故障发生时于测量时间窗口内注入攻击向量的能力。

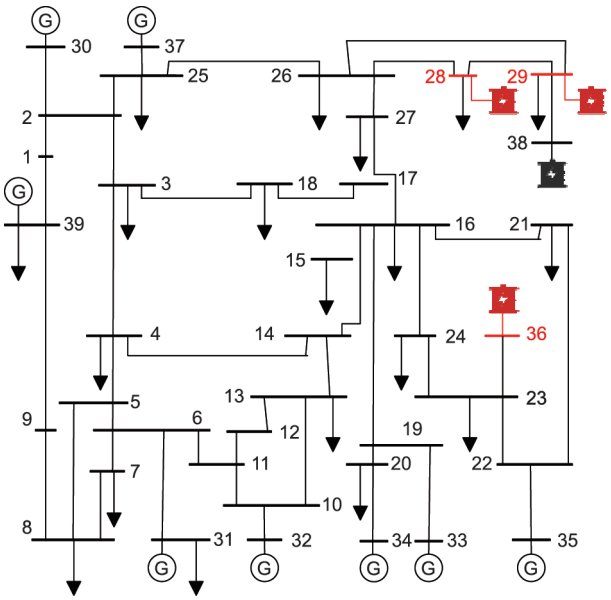

图5. 新英格兰39节点系统拓扑图。

5.2. 参数估计

设置采样间隔为1 ms，数据点总数N为5000。在模拟分布式电源稳态运行的过程中，攻击者监测u、i、u_g、i_g和u^*的测量数据，其中含有高斯噪声$\forall\xi_i\in\xi$，$\xi_i\sim N\left(0,0.01^2\right)$。采用步骤2的方法，对LCL滤波器和DC/AC控制器参数进行估计，如表1所示，所提算法能够准确估计各参数。

5.3. 攻击向量生成

将攻击目标的采样间隔设置为400 μs，在时间窗口内共采集250个数据点。据此，攻击阶段数α确定为5，每个阶段采集10个数据点。GAN训练完成后，随机选取50个样本用于评估攻击有效性。将GAN生成的攻击向量注入目标后，其影响结果如图6所示，其中0代表不稳定，1代表稳定。攻击实施后，预测精度从98.75%下降至56.00%，所有成功篡改的样本均将预测结果从稳定误导为不稳定。这一结果是由于训练集中标记为稳定的样本数量显著多于标记为不稳定的样本，导致模型主要学习稳定状态的特征。因此，攻击者更容易将预测结果从稳定误导为不稳定。

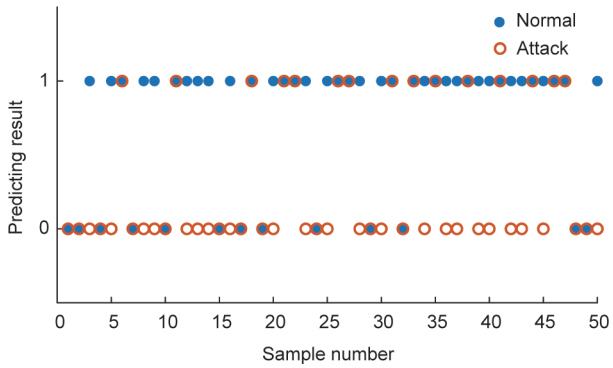

图6. 攻击向量注入前后的预测结果。

以成功欺骗攻击目标的样本为例，图7展示了拟注入节点28、29和36测量数据的攻击向量$u_{gd,qT}$。该攻击方法的总计算时间约为750 μs，不足两个采样间隔。因此，$h=11$，即前11个数据点未被攻击者篡改。注入攻击向量后，尽管仿真结果证实电力系统稳定，目标的预测结果却从稳定转变为不稳定。

根据步骤4的方法，计算得到每个分布式电源内DC/AC变换器的最优输出序列$u^*_{d,qA}$，如图8所示。随后在步骤5中，计算篡改的测量序列$u_{d,qA}$和$i_{d,qA}$，如图9所示。

表1 LCL滤波器和DC/AC变换器控制器参数的真实值与估计值

Value	LCL filter parameter					DC/AC converter controller parameter							
	R_f (Ω)	L_f (H)	C_f (F)	R_g (Ω)	ω	K_{Pi}	K_{Ii}	K_{Pu}	K_{Iu}	$K_{\omega C}$	$K_{\omega L}$ (kV)	U_{dr} (kV)	U_{qr} (kV)
True	0	0.006	1.000×10^{-5}	0	120π	0.8424	0.1677	1.0604	0.1084	0.0042	0.0486	17.560	4.460
Estimated	4.795×10^{-4}	0.006	0.921×10^{-5}	7.571×10^{-6}	375.7149	0.8421	0.1684	1.0608	0.1060	0.0042	0.0487	17.213	4.589

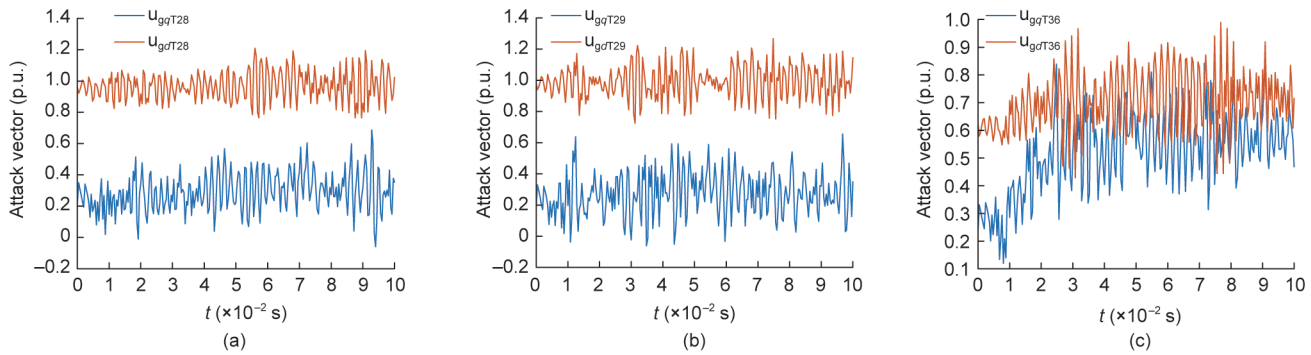

图7. GAN生成攻击向量。（a）注入节点28的攻击向量；（b）注入节点29的攻击向量；（c）注入节点36的攻击向量。p.u.：标幺值。

通过适当调整式（43）中 W_u 和 W_i 的值，可使 $i_{d,qA}$ 在每个阶段保持几乎恒定。

将篡改序列 $u_{d,qA}$ 和 $i_{d,qA}$ 注入DC/AC控制器后，输出 $u_{d,qA}^{*\prime}$ 与最优序列 $u_{d,qA}^{*}$ 的对比如图10所示。其中，r 表示 $u_{d,qA}^{*}$ 与 $u_{d,qA}^{*\prime}$ 之间的差异，计算方法如下所示。由于参数估计存在误差，随着篡改测量序列计算的迭代累积，$u_{d,qA}^{*}$ 和 $u_{d,qA}^{*\prime}$ 在每个阶段后半部分出现差异，而在每个阶段的前半部分差异较小。因此，可通过增加攻击阶段数量 α 来提高攻击精度。

$$r=\log_{10}\frac{\left|u_{dA}^{*}-u_{dA}^{*\prime}\right|+\left|u_{qA}^{*}-u_{qA}^{*\prime}\right|}{2} \tag{55}$$

在新英格兰39节点系统中，入侵三个分布式电源的攻击成功率为43.29%。与直接向通信网络注入攻击向量（文献[41]中的成功率为70.67%）相比，尽管成功率下降，但攻击者的入侵难度大大降低，攻击者只需要向分布式电源注入扰动，而无需向防护严密的通信网络注入攻击向量。

5.4. 针对不同攻击目标的攻击效果对比

为了评估所提攻击方法在不同架构神经网络中的有效性，本研究将先前生成的攻击向量应用于VGG16、SqueezeNet、ResNet18和InceptionV3模型中。表2汇总了这些模型在正常和受攻击条件下的暂态稳定性预测精度。

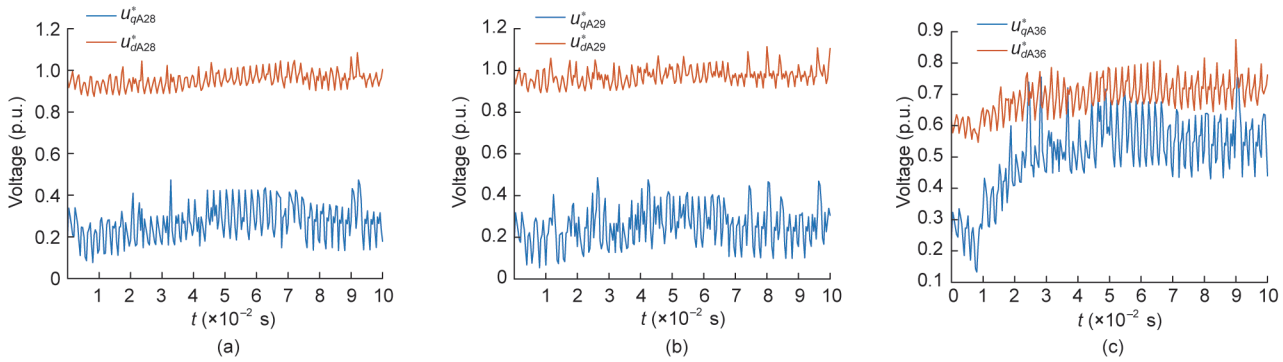

图8. DC/AC变换器的最优输出序列。（a）节点28上变换器的输出；（b）节点29上变换器的输出；（c）节点36上变换器的输出。

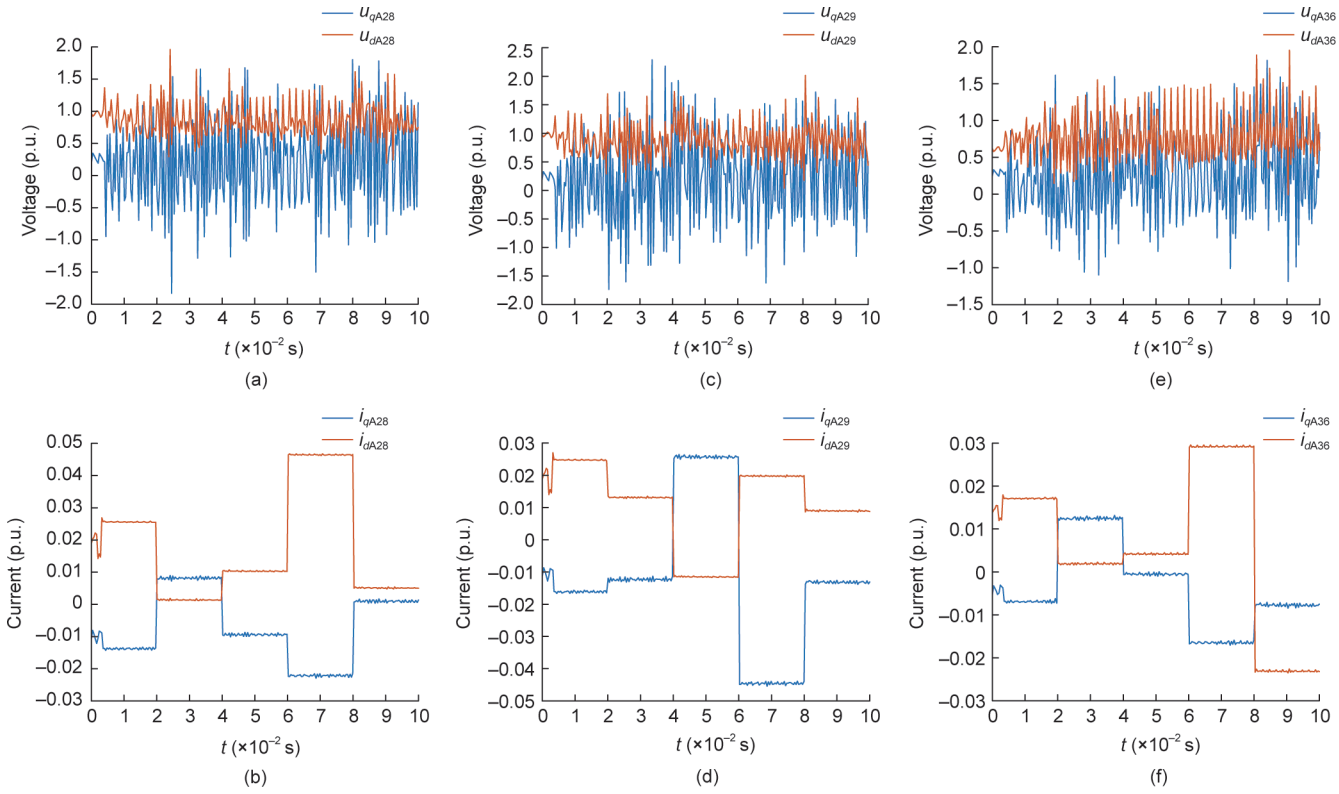

图9. DC/AC控制器的篡改测量序列。（a）、（c）、（e）分别为节点28、29、36上控制器的篡改电压测量序列；（b）、（d）、（f）分别为节点28、29、36上控制器的篡改电流测量序列。

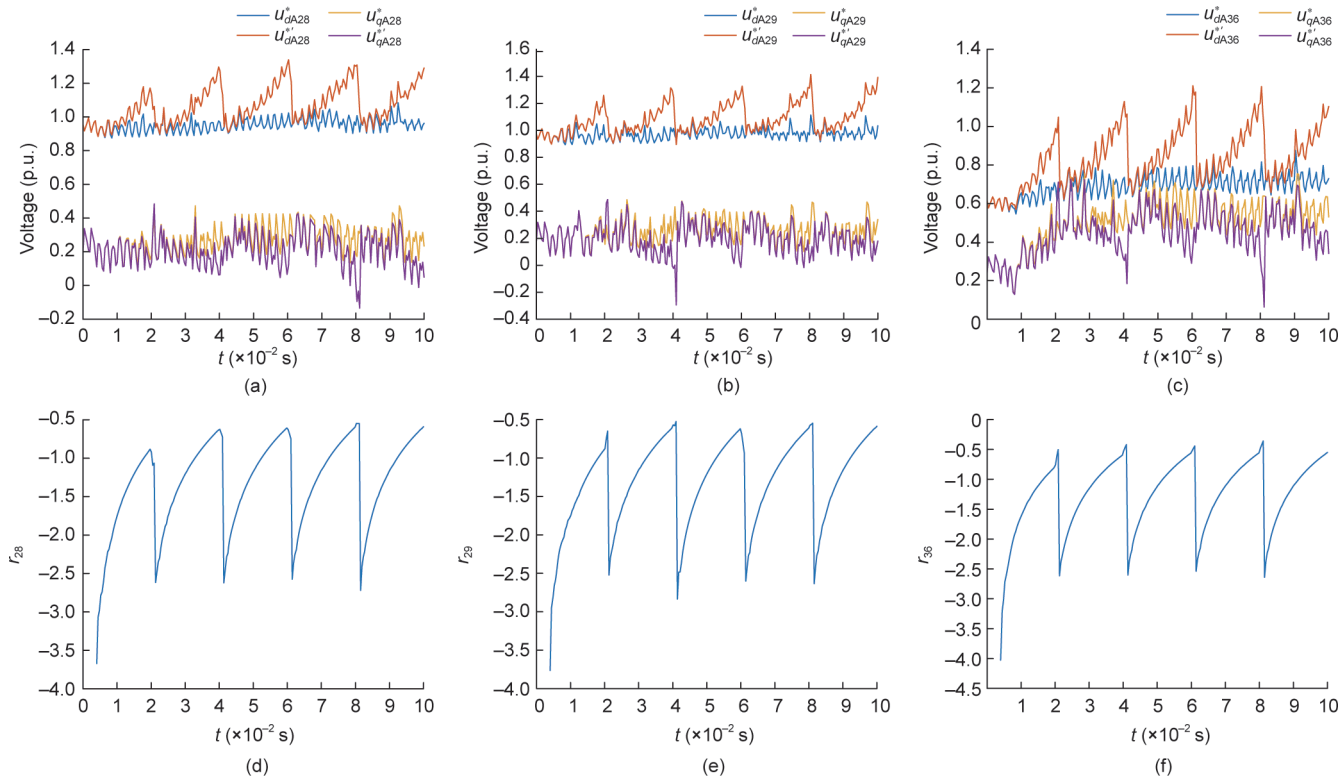

图10. $u^{*}_{d,qA}$ 与 $u^{*'}_{d,qA}$ 数据对比，其中 $u^{*}_{d,qA}$ 是DC/AC变换器的最优输出序列，$u^{*'}_{d,qA}$ 是注入篡改测量数据后的输出序列。（a）、（c）、（e）分别是节点28、29、36上DC/AC变换器的输出序列；（b）、（d）、（f）是 $u^{*}_{d,qA}$ 与 $u^{*'}_{d,qA}$ 之间的差异。

表2 针对不同攻击目标的攻击效果对比

Neural network used to predict transient stability	Number of layers	Accuracy		Success rate
		Normal	Attack	
AlexNet	23	98.75%	56%	43.29%
VGG16	41	98.75%	40%	59.49%
SquezzeNet	69	98.25%	58%	40.97%
ResNet18	71	99.25%	72%	27.46%
InceptionV3	315	98.50%	74%	24.87%

结果表明，攻击实施后，模型的预测精度均显著下降。与AlexNet和VGG16的简单架构相比，ResNet18和InceptionV3由于具有更精确的判别边界、更多的网络层和更复杂的结构，表现出较低的脆弱性。这一结果证实了本方法生成攻击向量的可迁移性，即无论攻击者是否掌握特定神经网络架构的信息，攻击方法均能有效欺骗目标。

为了验证所提攻击方法的泛化能力，本文分别在IEEE-57节点、118节点和145节点系统上开展实验验证，这些系统可代表不同规模和不同复杂程度的典型电网。负载变化范围设定为基准值的80%~120%，以模拟电网需求的实际波动情况。本研究在每个系统内任意线路的随机位置引入三相短路故障，故障持续时间在0.1~0.35 s之间变化。为全面模拟各种运行模式和故障场景，实验采用蒙特卡洛方法为每个系统生成2000组样本。

针对不同规模系统的攻击效果如表3所示，正常运行条件下，各系统的TSP均可保持较高的预测精度，当注入构造的攻击向量后，预测精度显著下降，证明了本文所提攻击方法的有效性。在较大规模系统（IEEE 118节点和145节点系统）中，精度下降尤为明显，凸显了本文所述攻击方法的可扩展性。

表3 不同规模电力系统中的攻击效果

The attacked power system	The buses of the attacked power supplies	Accuracy		Success rate
		Normal	Attack	
IEEE 57-bus system	4, 15, 38, 43	98.40%	48%	51.22%
IEEE 118-bus system	12, 19, 36, 59, 76, 99	97.35%	62%	36.31%
IEEE 145-bus system	22, 43, 61, 79, 101, 109	97.10%	66%	32.03%

上述实验结果表明，本文所提攻击方法可在不同规模系统中对数据驱动算法构成严重威胁，而且可以有效地推广至不同电网环境中。这些结果凸显了开发更强大防御机

制的迫切需求，此类机制需能随智能电网规模的不断扩大和复杂性的日益提升而相应扩展。

6. 结论

针对网络攻击的研究可丰富异常样本库，为数据驱动控制策略的设计和训练提供支撑，从而增强系统对攻击的防御能力。本文提出了一种针对电力系统数据驱动算法的攻击方法，设计并训练了基于GAN的攻击向量生成器，并通过分布式电源将攻击向量实时注入电力系统。针对TSP策略的攻击验证了方法的有效性，针对不同规模系统的攻击验证了方法的可迁移性。

CRediT authorship contribution statement

Zengji Liu: Conceptualization, Data curation, Formal Analysis, Methodology, Funding acquisition, Project administration, Resources, Supervision, Software, Validation, Visualization, Writing-original draft, Writing-review & editing. **Mengge Liu:** Investigation, Writing-original draft, Writing-review & editing. **Qi Wang:** Writing-original draft, Writing-review & editing. **Yi Tang:** Writing-original draft, Writing-review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

致谢

本研究由国家自然科学基金（62302234）资助。

References

[1] Xia X, Lin J, Jia Q, Wang X, Ma C, Cui J, et al. ETD-ConvLSTM: a deep learning approach for electricity theft detection in smart grids. IEEE Trans Informn Forensics Secur 2023;18:2553–68.
[2] Ruan J, Fan G, Zhu Y, Liang G, Zhao J, Wen F, et al. Super-resolution perception assisted spatiotemporal graph deep learning against false data injection attacks in smart grid. IEEE Trans Smart Grid 2023;14(5):4035–46.
[3] Cao D, Zhao J, Hu W, Yu N, Hu J, Chen Z. Physics-informed graphical learning and bayesian averaging for robust distribution state estimation. IEEE Trans Power Syst 2023;39(2):2879–92.
[4] An D, Zhang F, Cui F, Yang Q. Toward data integrity attacks against distributed dynamic state estimation in smart grid. IEEE Trans Autom Sci Eng 2023;21(1):881–94.
[5] Li B, Xu Q. A machine learning-assisted distributed optimization method for inverter-based Volt-VAR control in active distribution networks. IEEE Trans Power Syst 2023;39(2):2668–81.
[6] Gao Y, Ai Q, He X, Fan S. Coordination for regional integrated energy system through target cascade optimization. Energy 2023;276:127606.
[7] Rojek I, Mikołajewski D, Mroziński A, Macko M. Machine learning- and artificial intelligence-derived prediction for home smart energy systems with pv installation and battery energy storage. Energies 2023;16(18):6613.
[8] Xiao Z, Tang F, Wang M. Wind power short-term forecasting method based on lstm and multiple error correction. Sustainability 2023;15(4):3798.
[9] Liu Z, Li X, Zhao H. Short-term wind power forecasting based on feature analysis and error correction. Energies 2023;16(10):4249.
[10] Wang G, Wang C, Shahidehpour M, Lin W. Deep semi-supervised learning method for false data detection against forgery and concealing of faults in cyber–physical power systems. IEEE Trans Smart Grid 2023;15(1):944–58.
[11] Mishra DK, Wang J, Li L, Zhang J, Hossain MJ. Resilience-driven scheme in multiple microgrids with secure transactive energy system framework. IEEE Trans Ind Appl 2023;60(2):2277–89.
[12] Yang T, Yan B, Yang Q, Wang W. A packet payload anomaly detection approach for cyber–physical power systems. CSEE J Power Energy Syst. In Press.
[13] Wang Z, Qi D, Li Z, Zhang J. Valid false data injection attacks under approximated power flow model in power distribution systems. CSEE J Power Energy Syst. In Press.
[14] Ruan J, Wang Q, Chen S, Lyu H, Liang G, Zhao J, et al. On vulnerability of renewable energy forecasting: adversarial learning attacks. IEEE Trans Ind Inform 2023;20(3):3650–63.
[15] Zhao T, Yue M, Wang J. Robust power system stability assessment against adversarial machine learning-based cyberattacks via online purification. IEEE Trans Power Syst 2023;38(6):5613–22.
[16] Sahu A, Davis K. Inferring adversarial behaviour in cyber–physical power systems using a Bayesian attack graph approach. IET Cyber Phys Syst Theory Appl 2023;8(2):91–108.
[17] Sarker IH, Janicke H, Ferrag MA, Abuadbba A. Multi-aspect rule-based AI: methods, taxonomy, challenges and directions toward automation, intelligence and transparent cybersecurity modeling for critical infrastructures. Int Things 2024;25:101110.
[18] Sahani N, Zhu R, Cho JH, Liu CC. Machine learning-based intrusion detection for smart grid computing: a survey. ACM Trans Cyber Phys Syst 2023; 7(2):1–31.
[19] Diaba SY, Elmusrati M. Proposed algorithm for smart grid DDoS detection based on deep learning. Neural Networks 2023;159:175–84.
[20] Radhoush S, Vannoy T, Liyanage K, Whitaker BM, Nehrir H. Distribution system state estimation and false data injection attack detection with a multi-output deep neural network. Energies 2023;16(5):2288.
[21] Yan B, Jiang Z, Yao P, Yang Q, Li W, Zomaya AY. Game theory based optimal defensive resources allocation with incomplete information in cyber–physical power systems against false data injection attacks. Prot Contr Mod Pow 2024;9(2):115–27.
[22] Reda HT, Anwar A, Mahmood AN, Tari Z. A axonomy of cyber defence strategies against false data attacks in smart grids. ACM Comput Surv 2023;55(14):1–37.
[23] Liu X, Liu J, Zhao Y, Ding T, Liu X, Liu J. A Bayesian deep learning-based probabilistic risk assessment and early-warning model for power systems considering meteorological conditions. IEEE Trans Ind Inform 2023;20(2):1516–27.
[24] Zhang Z, Hu J, Lu J, Cao J, Yu J. False data injection attacks on LFC systems: an AI-based detection and countermeasure strategy. IEEE Trans Circuits Syst I 2023;71(5):1969–77.
[25] Zhang Q, Li F, Zhao J, She B. False data injection attack and corresponding countermeasure in multienergy systems. IEEE Trans Power Syst 2023; 39(2):3537–47.
[26] Ge X, Yue M. A detection strategy based on deep learning against sequential outages induced by false data injection attacks. Electr Eng 2024;106:5201–17.
[27] Wei S, Xu J, Wu Z, Hu Q, Yu X. A false data injection attack detection strategy for unbalanced distribution networks state estimation. IEEE Trans Smart Grid 2023;14(5):3992–4006.
[28] Ye X. Sparsity and coordination constraints on stealth data injection attacks [dissertation]. Sheffield: University of Sheffield; 2023.
[29] Kong M, Shen F, Du P, Peng X, Zhong W. Distributed secure consensus for multiagent systems based on removing intra-cluster coupling restrictions and its

application to energy systems. Inform Sci 2024;653:119579.
[30] Ahmadian S, Malki H, Han Z. Cyber attacks on smart energy grids using generative adverserial networks. In: Proceedings of 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 2018 Nov 26–29; Anaheim, CA, USA. IEEE; 2018.
[31] Efatinasab E, Brighente A, Rampazzo M, Azadi N, Conti M. GAN-GRID: a novel generative attack on smart grid stability prediction. 2024. arXiv: 2405.12076.
[32] Dash SP, Khandeparkar KV. A false data injection attack on data-driven strategies in smart grid using GAN. Advances and Trends in Artificial Intelligence. Theory and applications. Cham: Springer; 2023.
[33] Ren C, Xu Y. Robustness verification for machine learning-based power system dynamic security assessment models under adversarial examples. IEEE Trans Control Netw Syst 2022;9(4):1645–54.
[34] Ren C, Du X, Xu Y, Song Q, Liu Y, Tan R. Vulnerability analysis, robustness verification, and mitigation strategy for machine learning-based power system stability assessment model under adversarial examples. IEEE Trans Smart Grid 2021;13(2):1622–32.
[35] Cui L, Guo L, Gao L, Cai B, Qu Y, Zhou Y, et al. A covert electricity-theft cyber-attack against machine learning-based detection models. IEEE Trans Ind Inform 2022;18(11):7824–33.
[36] Khan OGM, El-Saadany EF, Youssef A, Shaaban MF. Cyber security of market-based congestion management methods in power distribution systems. IEEE Trans Ind Inform 2021;17(12):8142–53.
[37] Kundu A, Sahu A, Serpedin E, Davis K. A3D: Attention-based auto-encoder anomaly detector for false data injection attacks. Electr Power Syst Res 2020; 189:106795.
[38] Takiddin A, Ismail M, Zafar U, Serpedin E. Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids. IEEE Syst J 2022; 16(3):4106–17.
[39] Boyaci O, Umunnakwe A, Sahu A, Narimani MR, Ismail M, Davis KR, et al. Graph neural networks based detection of stealth false data injection attacks in smart grids. IEEE Syst J 2021;16(2):2946–57.
[40] Boyaci O, Narimani MR, Davis KR, Ismail M, Overbye TJ, Serpedin E. Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks. IEEE Trans Smart Grid 2021;13(1):807–19.
[41] Liu Z, Wang Q, Ye Y, Tang Y. A GAN based data injection attack method on data-driven strategies in power systems. IEEE Trans Smart Grid 2022; 13(4): 3203–13.