Modeling and Simulation 建模与仿真, 2023, 12(3), 2182-2196
Published Online May 2023 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2023.123200

Hans汉斯

基于WGAN状态重构的智能电网虚假数据注入攻击检测

张　笑，孙　越

上海理工大学机械工程学院，上海

收稿日期：2023年3月1日；录用日期：2023年5月9日；发布日期：2023年5月16日

摘　要

针对虚假数据定位检测适应性低、篡改量测影响系统状态精确感知的问题，提出一种基于WGAN (Wasserstein generative adversarial networks, WGAN)状态重构的智能电网虚假数据检测与修正模型。首先，根据历史状态变量的概率分布，初步锁定并剔除具有潜在攻击风险的状态变量。然后，采用Wasserstein生成对抗网络重构缺失变量，WGAN通过Wasserstein距离衡量生成分布与真实分布之间的差异，能够生成有意义的梯度以优化网络模型参数。最后，以重构状态作为一种状态参考，精确定位攻击节点，并结合网络拓扑参数修正篡改量测值。将纽约州数据用在IEEE-14节点测试系统，进一步验证所提方法的可行性与有效性。

关键词

虚假数据，篡改量测，状态估计，状态重构，生成对抗网络

Detection of False Data Injection Attack in Smart Grid Based on WGAN State Reconfiguration

Xiao Zhang, Yue Sun

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 1st, 2023; accepted: May 9th, 2023; published: May 16th, 2023

Abstract

Aiming at the problems of low adaptability of false data location detection and the influence of

文章引用: 张笑, 孙越. 基于 WGAN 状态重构的智能电网虚假数据注入攻击检测[J]. 建模与仿真, 2023, 12(3): 2182-2196.
DOI: 10.12677/mos.2023.123200

tamper measurement on the accurate state awareness for the power system, a false data detection and correction model of smart grid based on state reconstruction of Wasserstein Generative Adversarial Networks (WGAN) was proposed. Firstly, the state variables with potential attack risk are initially locked and eliminated according to the probability distribution of historical state variables. Then, the missing state variables are reconstructed by Wasserstein generative adversarial networks. WGAN measures the difference between the generated distribution and the real distribution through Wasserstein distance, which can generate meaningful gradients to optimize the parameters of the network model. Finally, the reconstructed variables are used as a state reference to locate the attacked bus, and to correct the measurement data combined with the network topology parameters. The feasibility and validity of the proposed method are further verified in IEEE-14 bus test system with the New York data.

Keywords

False Data, Tamper Measurement, State Estimation, State Reconstruction, Generative Adversarial Network

1. 引言

随着电力物理系统与信息系统深度耦合，智能电网可观测性显著增强，但也面临网络安全的威胁[1] [2]。虚假数据注入攻击(false data injection attacks, FDIA)作为智能电网中兼具破坏性与隐蔽性的网络攻击[3]，通过篡改电力量测数据，使状态估计失去对系统状态的准确感知，引起能源管理系统(energy management system, EMS)做出错误决策，进而破坏电力系统安全稳定运行[4]。因此，研究高效 FDIA 检测与修正方法对保障智能电网安全性、可观测性具有重要意义。

现有文献针对 FDIA 的防范研究主要集中在物理防御与在线检测两个方面。物理防御专注加密一组现有量测或优化配置向量测量单元(phasor measurement unit, PMU)，防范电力量测数据遭到篡改。文献[5]通过图形分析法加密最优量测子集，遏制 FDIA。文献[6]以 PMU 绝对安全为前提，采用贪婪算法确定 PMU 最佳位置防范 FDIA。然而，文献[7]指出物理防御无法确保加密量测或 PMU 绝对安全，加密设备的高额资金也限制了在大型电网中实际应用。为此，基于数据的在线检测方法被提出并持续改进。统计学方法最先用于 FDIA 检测，文献[8]通过计算攻击前后量测变化概率分布之间的 KLD (Kullback-Leibler divergence)进行检测，但面对数据的激增、攻击方式的演变呈现局限性。而深度机器学习具有出色的从大数据中提取特征的能力，文献[9]利用改进的卷积神经网络，识别虚假数据的行为特征，提高分类检测精度。

上述研究集中于检测攻击存在与否，根据实际电力系统的安全要求，需进一步识别攻击节点与受损量测，为 EMS 操作人员隔离攻击节点、执行量测物理保护提供必要的数据支撑。文献[10]采用基于机器学习的一类一网检测模型为小节点系统每个节点构建分类检测器，在节点数量众多的大型电力系统复杂性增加。文献[11]将发电机的输出功率视为绝对安全序列，通过维特比算法预测状态值作为参考，识别攻击节点。这种方法的适用性增强，但对低强度的节点状态值攻击，检测率不理想。为进一步提高攻击节点检测精度，文献[12]基于自适应无迹卡尔曼滤波算法估计系统状态，经一致性检验、虚假数据隐蔽性检

验，提升识别率。估计算法假设噪声符合正态分布、量测误差为常数阵，与实际电网噪声时变性、量测误差在不同节点、不同线路存在差异性不相符，影响状态预测精度，进而影响攻击识别率。

生成对抗网络(generative adversarial networks, GAN)因能充分学习复杂数据之间的分布规律，已被用于电网不同类型量测缺失重构工作。文献[13]利用无监督 GAN 自主提取数据特征，并结合双重语义感知重构约束，使高比例有功量测缺失重构精度达 100%。文献[14]以 Wasserstein 距离代替 JS 散度，采用 WGAN 克服了 GAN 训练过程导致的梯度消失问题，通过真实性约束、上下文约束挑选最优生成量测，随机缺失量测重构误差低于 7.0 e^{-4}pu。而高精度随机数据缺失重构的实现，契合局部网络拓扑结构下随机节点状态受到攻击的问题，为从状态重构角度建模，解决定位检测模型中状态预测精度低、节点分类器适应性受限的问题，提供了理论依据。

因此，本文利用 WGAN 重构精度高的优势，提出一种基于状态变量重构的智能电网虚假数据检测与修正模型。根据历史状态变量的概率分布，采用“3σ”准则锁定并剔除具有潜在攻击风险的状态变量；WGAN 通过 Wasserstein 距离对状态变量进行无监督重构训练，使生成器模型具备快速重构出接近真实状态值的能力；将生成器重构的状态值作为一种状态参考，定位攻击节点并替换受损状态，结合网络拓扑参数修正篡改量测值。算例仿真验证所提方法的有效性。

2. 虚假数据基本原理

2.1. 状态估计与不良数据检测

在电网中，控制中心通过状态估计，利用实时量测系统的冗余性提高数据精度，自动消除随机干扰引起的误差信息[15]。电力系统状态估计是通过各种智能仪表的监测数据，对当前系统运行状态评估的过程。其结果常用于经济调度、安全评估等。在交流状态估计模型中，给定量测数据 $\boldsymbol{z}$，通过求解最小化目标函数，计算系统状态值 $\boldsymbol{x}$。

$$J(\boldsymbol{x})=\left[\boldsymbol{z}-h(\boldsymbol{x})\right]^{\mathrm{T}}\boldsymbol{R}^{-1}\left[\boldsymbol{z}-h(\boldsymbol{x})\right] \tag{1}$$

式中，$\boldsymbol{z}$ 为包含电压幅值、节点注入功率、线路传输功率的量测向量；$\boldsymbol{x}$ 为包含节点电压幅值、电压相角的系统状态变量；$h(\boldsymbol{x})$为量测向量与状态变量的非线性函数；$\boldsymbol{R}$ 为已知量测误差方差阵。

为发现因通信干扰、仪表故障可能导致的量测异常，传统不良数据检测机制通常基于残差 $\boldsymbol{r}$ 欧几里得范数 $\|\boldsymbol{r}\|_2$ 与预定检测阈值 τ 比较，大于阈值 τ 表明存在不良数据。残差定义为：

$$\boldsymbol{r}=\boldsymbol{z}-h(\boldsymbol{x}) \tag{2}$$

2.2. 局部虚假数据注入攻击数学模型

攻击者在掌握电力系统网络拓扑参数并能同时篡改多个量测的情况下[16]，设攻击受损量测为 $\boldsymbol{z}_a$，攻击后状态变量为 $\boldsymbol{x}_a$。此时，残差公式为：

$$\begin{aligned} r_a &= \left\|\boldsymbol{z}_a-h(\boldsymbol{x}_a)\right\|_2 \\ &= \left\|\boldsymbol{z}+\boldsymbol{a}-h(\boldsymbol{x}+\boldsymbol{c})\right\|_2 \\ &\le \left\|\boldsymbol{z}-h(\boldsymbol{x})\right\|_2+\left\|\boldsymbol{a}-h(\boldsymbol{x}+\boldsymbol{c})+h(\boldsymbol{x})\right\|_2 \\ &= \left\|\boldsymbol{r}\right\|_2+\tau_a \end{aligned} \tag{3}$$

式中，$\boldsymbol{a}$ 为攻击向量；$\boldsymbol{c}$ 为由攻击后状态偏差向量；τ_a 为由攻击引起的残差增量值。

由公式(4)知，当攻击向量满足

$$\boldsymbol{a}=h\left(\boldsymbol{x}+\boldsymbol{c}\right)-h\left(\boldsymbol{x}\right) \tag{4}$$

由攻击引起残差增量 τ_a 为 0，使得 $r_a=\left\|\boldsymbol{r}\right\|_2$，即可实现攻击的隐蔽性。

由公式(4)可知，攻击者掌握电网全部拓扑信息下的攻击成功率更高，但掌握局部信息，使局部节点状态受损的攻击更符合实际[17]。针对局部区域节点 i 的电压幅值攻击为例，局部攻击向量 $\boldsymbol{a}$ 中节点有功注入增量 ΔP_{ia}、节点无功注入增量 ΔQ_{ia} 为：

$$\Delta P_{ia}=\sum_{j\in\Omega_i}V_j\left[G_{ij}\cos\theta_{ij}+B_{ij}\sin\theta_{ij}\right]\left[-V_i+\left(V_i+\Delta V_{ia}\right)\right] \tag{5}$$

$$\Delta Q_{ia}=\sum_{j\in\Omega_i}V_j\left[G_{ij}\sin\theta_{ij}-B_{ij}\cos\theta_{ij}\right]\left[-V_i+\left(V_i+\Delta V_{ia}\right)\right] \tag{6}$$

式中，V_i、V_j 分别为节点 i、节点 j 的电压幅值；θ_{ij} 为节点 i 与节点 j 的电压相角差；ΔV_{ia} 为节点 i 的电压幅值攻击增量；G_{ij}、B_{ij} 分别为节点 i 与节点 j 之间支路电导、电纳；Ω_i 为节点 i 与相连节点构成的电力节点集合。

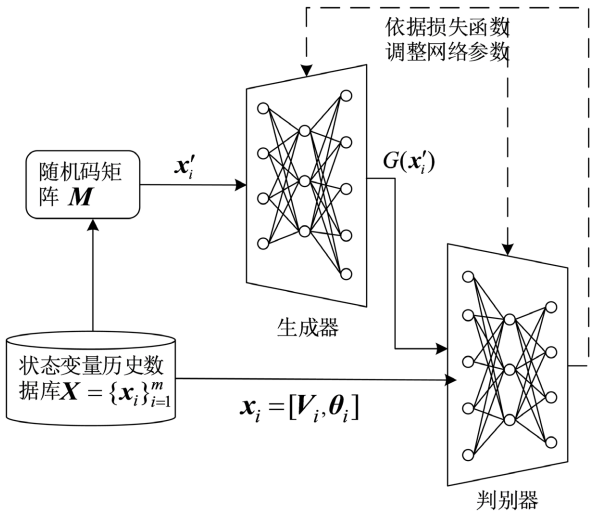
3. 基于 WGAN 状态重构的攻击检测模型

3.1. 基于 WGAN 的状态重构

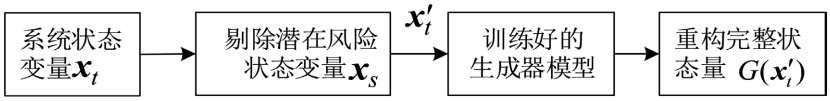
GAN 以 JS 散度作为损失函数衡量生成分布与真实分布的差异，面临两种分布没有重叠部分时，损失函数突变常值导致梯度消失。WGAN 以 Wasserstein 距离代替 JS 散度，即使两种分布没有重叠，Wasserstein 距离依旧平滑变化，生成有意义的梯度以优化网络模型参数，克服 GAN 训练过程中梯度消失问题[14]。

WGAN 由相互博弈的生成器与判别器两个对抗网络构成。从历史数据库选取 m 组无缺失状态变量构建数据集 $\boldsymbol{X}=\left\{\boldsymbol{x}_i\right\}_{i=1}^{m}$，样本数据 $\boldsymbol{x}_i=\left[\boldsymbol{V}_i,\boldsymbol{\theta}_i\right]$，设无缺失样本数据 $\boldsymbol{x}_i$ 之间复杂分布规律为 $p_r\left(\boldsymbol{x}_i\right)$。构建与样本数据 $\boldsymbol{x}_i$ 维度一致的随机二值掩码矩阵 $\boldsymbol{M}$，其中 0 代表随机缺失状态，1 代表未缺失部分。为改善重构效率，将样本数据 $\boldsymbol{x}_i$ 与 $\boldsymbol{M}$ 做哈达玛积(Hadamard product)运算，构成随机缺失 $\boldsymbol{x}_i'$ 作为生成器的输入[18]，通过生成器和判别器构建 $p_r\left(\boldsymbol{x}_i\right)$ 与重构状态 $p_g\left(\boldsymbol{x}_i'\right)$ 之间的映射关系。训练过程如图 1(a)所示。

重构检测阶段，针对 t 时刻状态变量 $\boldsymbol{x}_t$ 存在潜在攻击风险状态变量 $\boldsymbol{x}_s$ 情况，剔除 $\boldsymbol{x}_s$ 后，将缺失状态变量 $\boldsymbol{x}_t'$ 送入训练完成的生成器模型，重构完整状态量 $G\left(\boldsymbol{x}_t'\right)$，重构过程如图 1(b)所示。



(a) 模型训练



(b) 状态重构

Figure 1. State reconstruction basic structure based on WGAN
图 1. 基于 WGAN 的状态重构基本框架

3.2. WGAN 网络结构

生成器网络结构如表 1 所示，由 4 层卷积网络组成，不仅对状态变量进行特征提取，也确保输入与输出维度一致。卷积后添加批标准化层保证模型的非线性表达能力，使梯度传播层次更深[19]。卷积之后采用 Tanh 激活函数，输出 $14 \times 2 \times 1$ 的重构状态变量。

Table 1. Parameters of generator network of WGAN
表 1. WGAN 生成器网络参数

网络层	滤波器	卷积核	步长	填充	动量
卷积层 + ReLU	256	3	1	1	-
批标准化层	-	-	-	-	0.1
卷积层 + ReLU	64	3	1	1	-
批标准化层	-	-	-	-	0.1
卷积层 + ReLU	32	3	1	1	-
批标准化层	-	-	-	-	0.1
卷积层 + Tanh	1	3	1	1	-

判别器网络结构参数如表 2 所示。判别器输入为生成器生成样本 $G(x_i')$ 和真实样本 x_i'，网络结构类似于生成器的镜像。经过 3 层卷积特征提取后，Flatten 函数在不改变卷积层输出顺序的情况下将数据进行降为压平。最后由全连接层与 sigmoid 函数输出属于真实状态变量的概率。

Table 2. Parameters of discriminator network of WGAN
表 2. WGAN 判别器网络参数

网络层	滤波器	卷积核	步长	填充	动量
卷积层 + ReLU	32	3	1	1	-
批标准化层	-	-	-	-	0.1
卷积层 + ReLU	64	3	1	1	-
批标准化层	-	-	-	-	0.1
卷积层 + ReLU	128	3	1	1	-
批标准化层	-	-	-	-	0.1
Flatten 函数	-	-	-	-	-
全连接 + Sigmoid	-	-	-	-	-

3.4. 虚假数据检测与修正流程

FDIA 检测模型涉及潜在风险状态变量锁定、虚假数据隐蔽性检验，以及检测准确率指标。

1) 采用“3σ”准则锁定潜在攻击风险状态变量，公式如下：

$$\left|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\right|>\left|\boldsymbol{\mu}\pm 3\boldsymbol{\sigma}\right| \tag{7}$$

式中，$\boldsymbol{x}_t$ 为 t 时刻状态变量，$\boldsymbol{\mu}$、$\boldsymbol{\sigma}$ 分别为依据历史数据库计算状态变量变化量的均值、方差向量。

2) 虚假数据隐蔽性检验公式如下：

$$return\left(\left|\boldsymbol{x}_s-\boldsymbol{x}_r\right|\geq\tau_1\middle\|\tau_2\right) \tag{8}$$

式中，$\boldsymbol{x}_s$ 为锁定的潜在攻击风险状态变量；$\boldsymbol{x}_r$ 为生成器重构的攻击风险状态变量；τ_1、τ_2 分别为攻击检测电压幅值阈值、相角阈值；$return(\cdot)$ 为返回满足括号内条件向量元素的索引。

3) 采用检测准确率、查全率衡量模型性能，公式如下：

$$A_c=\frac{TN+TP}{TN+FN+TP+FP} \tag{9}$$

$$\eta_r=\frac{TP}{FN+TP} \tag{10}$$

式中，TN 为正常样本识别为正常的数量；FP 为正常样本识别攻击的数量；FN 为攻击样本识别为正常的数量；TP 为攻击样本识别为攻击的数量。

基于状态重构虚假数据检测与修正模型流程如图 2 所示，主要包含以下 4 个步骤。

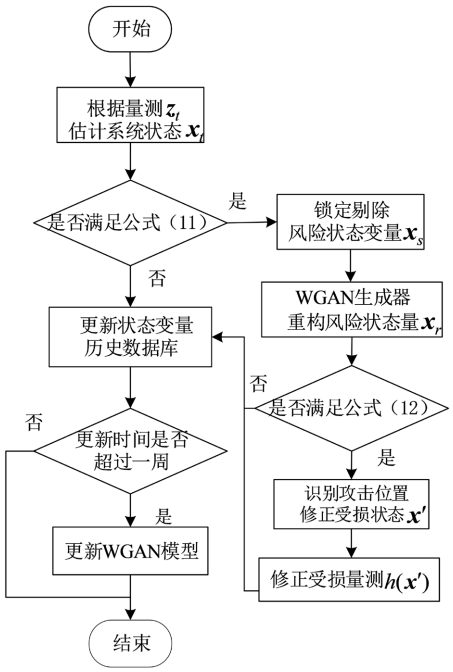


Figure 2. Flow chart of examine model against false data
图 2. 虚假数据检测与修正模型流程图

1) 根据系统获取的量测量，估计系统运行状态，采用公式(7)锁定是否具有潜在风险状态量。

2) 针对潜在风险状态变量，剔除后由 WGAN 生成器模型重构潜在风险状态估计值，并采用公式(8)定位识别节点是否受到攻击。

3) 受损状态值由重构值替换，构成状态修正量 $\boldsymbol{x}'$，结合网络拓扑参数修正受损量测 $h(\boldsymbol{x}')$。

4) 为充分学习状态变量随时间的变化规律，状态变量历史数据库保留一个月状态变量，每周更新一次基于 WGAN 状态重构的生成器模型。

4. 算例分析

在 IEEE14 节点系统中验证所用模型的有效性，系统网络拓扑结构如图 3 所示。MATLAB 和 MATPOWER 用于生成实验数据，WGAN 模型使用 python 语言在 pytorch 环境下搭建，PC 配置为 I5-8300H/8GB RAM。

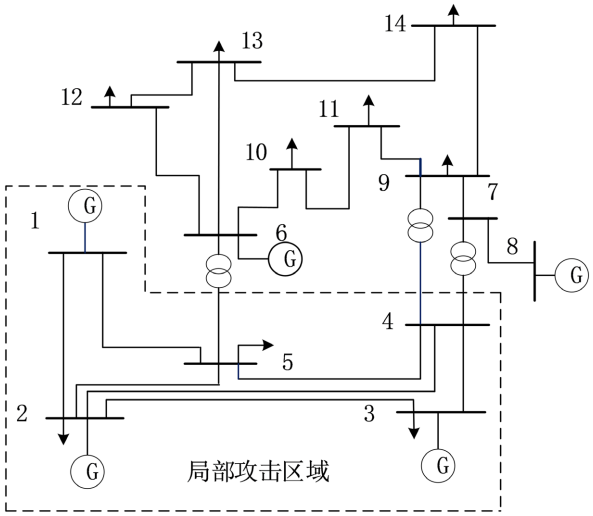


Figure 3. IEEE-14bus test system
图 3. IEEE-14 节点测试系统

4.1. 测试系统与模型训练参数

IEEE14 节点测试系统采用全量测配置方式，其量测向量 $\boldsymbol{z}$ 包含节点注入有功量测 14 个，节点注入无功量测 14 个，20 条支路有功量测 40 个、支路无功量测 40 个，节点电压量测 14 个，总计 112 个量测值。使用纽约独立系统运行商[20] (New York independent system operator, NYISO)2020 年 1、2 月份每 5 分钟的负荷数据模拟实际电力系统运行状况，状态数据的获取过程如下：

1) 将 IEEE14 节点系统的各个节点与纽约州 11 个地区的负荷匹配连接，匹配关系矩阵如下：

$$\begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 9 & 10 & 11 & 12 & 13 & 14 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{bmatrix}$$

矩阵第一行代表 IEEE-14 的节点编号，矩阵第二行为相应的 NYISO 地区编号。

2) 将纽约州 11 个地区负荷数据标幺化到与之对应的 IEEE-14 节点，初始化有功负荷和无功负荷，使测试系统运行在 IEEE-14 节点初始状态附近。由于 NYISO 没有提供无功负荷，假设测试系统负荷具有恒定的功率因数，计算得到节点无功负荷。

3) 假设发电机发电量与系统总负荷增速相同，计算新的系统总负荷与系统初始总负荷的比值，乘上系统所有发电机的出力。

4) 每隔五分钟通过 MATPOWER 潮流分析，将潮流计算值作为系统的量测真值，并在量测真值基础上添加均值为 0，标准差为 0.01 的高斯噪声，得到量测值。

5) 对收集的量测值送入控制中心，估计系统运行状态，并进行不良数据检测。

训练 WGAN 时，将历史数据库 1 月份前 30 天状态数据按 9:1 划分训练集 $\boldsymbol{X}_{train}$ 和测试集 $\boldsymbol{X}_{test}$，$\boldsymbol{X}_{train}$ 维度为(7776，14 × 2)，$\boldsymbol{X}_{test}$ 维度为(864，14 × 2)。设训练迭代次数 epochs 为 100，批处理量 batchsize 为 64，生成器学习率为 0.00005，判别器学习率为 0.00005。

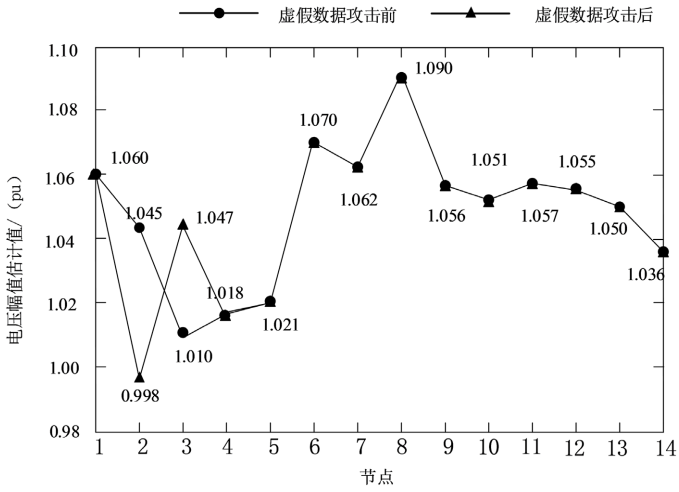
4.2. 局部虚假数据构建分析

为验证虚假数据构建的有效性，模拟 IEEE-14 节点系统的局部区域有限节点受到攻击的场景。假设攻击者掌握电力系统局部网络拓扑结构，并具备篡改局部区域相关量测的能力。以图 3 的攻击区域为例，假定攻击后节点 2 的电压幅值偏差增量 ΔV_{2a} 为−0.047，节点 3 的电压幅值偏差增量 ΔV_{3a} 为 0.037，基于式(4)构建针对 V_2、V_3 攻击向量 $\boldsymbol{a}$ 的非零元素值，局部网络拓扑下的量测参数的变化见表 3。

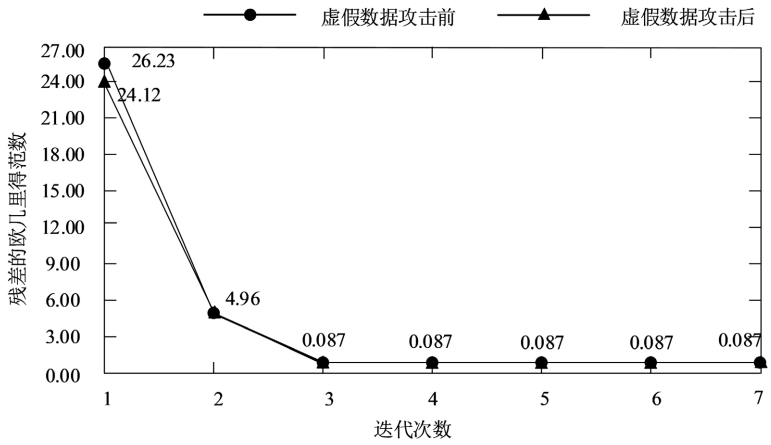
Table 3. IEEE-14 bus system local attack measurements
表 3. IEEE-14 节点系统局部攻击量测数据

量测位置	量测数据	真值	量测值	攻击增量
节点 1	P_1	2.323	2.325	0.182
	Q_1	−0.165	−0.164	0.779
节点 2	P_2	0.183	0.190	−0.473
	Q_2	0.309	0.312	−1.614
节点 3	P_3	−0.942	−0.945	0.173
	Q_3	0.061	0.059	0.608
节点 4	P_4	−0.478	−0.473	0.036
	Q_4	0.039	0.042	0.042
节点 5	P_5	−0.076	−0.078	0.098
	Q_5	−0.016	−0.024	0.243
支路 1-2	$P_{1\text{-}2}$	0.157	0.149	0.182
	$Q_{1\text{-}2}$	−0.204	−0.207	0.779
支路 1-5	$P_{1\text{-}5}$	0.755	0.762	0
	$Q_{1\text{-}5}$	0.039	0.046	0
支路 2-3	$P_{2\text{-}3}$	0.732	0.736	−0.104
	$Q_{2\text{-}3}$	0.036	0.029	−0.405
支路 2-4	$P_{2\text{-}4}$	0.561	0.548	−0.104
	$Q_{4\text{-}5}$	−0.016	−0.014	−0.238
支路 2-5	$P_{2\text{-}5}$	0.415	0.419	−0.098
	$Q_{2\text{-}5}$	0.012	0.018	−0.243
支路 3-4	$P_{3\text{-}4}$	−0.233	−0.235	0.068
	$Q_{3\text{-}4}$	0.045	0.054	0.198
支路 4-5	$P_{4\text{-}5}$	−0.612	−0.605	0
	$Q_{4\text{-}5}$	0.158	0.159	0

将攻击向量 $\boldsymbol{a}$ 叠加在量测向量 $\boldsymbol{z}$ 形成虚假数据 $\boldsymbol{z}_a$ 送入控制中心进行状态估计，图 4 为局部拓扑节点攻击前后电压幅值估计值、残差变化情况。



(a) 攻击前后节点电压幅值估计值对比



(b) 攻击前后残差的欧几里得范数变化

Figure 4. IEEE-14bus system local false data attack effect
图 4. IEEE-14 节点系统局部虚假数据攻击结果

由图 4 可见：

1) V_2 由原来的 1.045 降低为 0.998，V_3 由原来的 1.010 上升为 1.047，而系统的残差值经多次迭代依旧为 0.087，成功躲避不良数据检测机制，实现了局部攻击的隐蔽性。

2) 攻击者在掌握局部网络拓扑结构情况下，构建的电压幅值偏差增量 ΔV_{2a}、ΔV_{3a} 能准确干扰 V_2、V_3 到特定值，而不影响其他状态变量的变化。ΔV_{2a}、ΔV_{3a} 的出现打破原状态变量的内部相关性，为从状态重构角度出发，实现定位检测，提供了理论依据。

4.3. 重构效果分析

WGAN 模型训练完成后，在测试集中添加随机掩码矩阵 $\boldsymbol{N}$，设 $\boldsymbol{N}$ 中 0 代表随机缺失状态变量，1 代表未缺失部分。在 IEEE14 节点系统中，随机设定缺失节点状态数据送入生成器模型中进行重构，测试系统节点状态数据在单节点缺失、多节点缺失情况下的绝对误差均值与重构平均时间。表 4 给出单节点状

态缺失情况下不同节点电压幅值与相角的绝对误差均值。图 5 为多节点缺失情况下的绝对误差均值与重构平均时间。

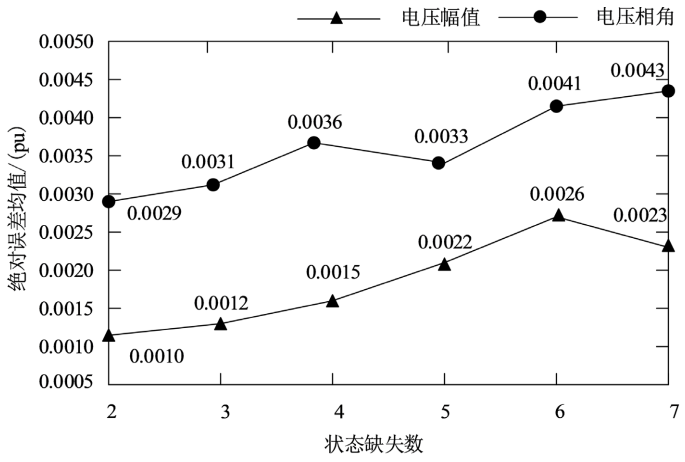
Table 4. Reconstruction result of missing state of single node based on WGAN model
表 4. 基于 WGAN 模型单节点缺失状态重构效果

节点序号	幅值绝对误差均值/(pu)	相角绝对误差均值/(pu)
1	0.00087	0
2	0.00095	0.00271
3	0.00114	0.00243
4	0.00094	0.00314
5	0.00072	0.00296
6	0.00109	0.00328
7	0.00064	0.00278
8	0.00138	0.00347
9	0.00075	0.00295
10	0.00108	0.00339
11	0.00083	0.00248
12	0.00102	0.00326
13	0.00127	0.00272
14	0.00099	0.00317

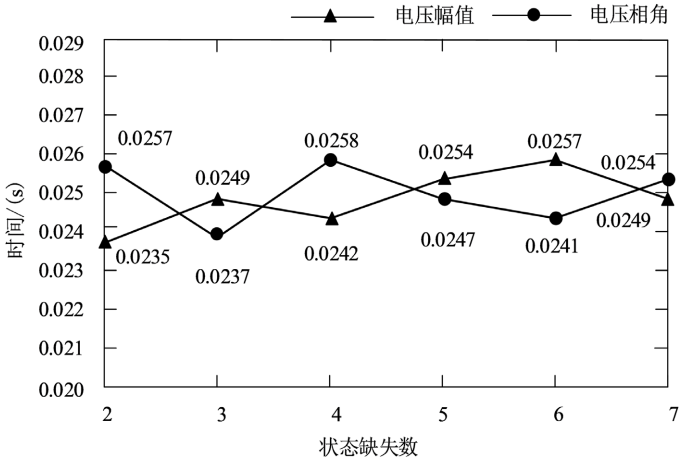
由表 4 可见：

1) 单节点状态缺失情况下，电压幅值与相角重构误差低于量测误差，WGAN 模型可为系统提供准确的状态值。

2) 电压幅值的重构效果要优于相角，其原因是负荷变化时，电力系统电压幅值的波动幅度相比于相角更加稳定。



(a) 多节点状态缺失的绝对误差均值



(a) 多节点状态缺失的重构平均时间

Figure 5. WGAN multi-node missing reconstruction result
图 5. WGAN 多节点缺失重构效果

从图 5 中可见：

1) 当状态缺失数小于等于 4 时，电压幅值与相角的重构绝对误差均值波动较小，具有较为稳定的重构效果；随着缺失量的进一步增加，电压幅值与相角的绝对误差均值增长较快，但重构电压幅值绝对误差均值保持在 0.003 之下，重构相角绝对误差均值保持在 0.005 之下，WGAN 模型具有较好的重构效果。

2) 不同缺失比列对状态重构的平均时间影响较小，重构平均时间在 0.024s 左右，重构效率得到保证。

3) 在状态缺失比列不断增大情况下，WGAN 模型依旧能充分学习状态变量之间时空特征、相互性、周期性等变化规律，具备快速重构出接近真实状态值的能力。

4.4. 检测阈值设定

式(9)中攻击检测电压阈值 τ_1、相角阈值 τ_2 选取的合理性直接影响检测结果的准确性。若阈值设定过低，未被攻击的节点将被判定为攻击；若阈值选取过高，攻击节点漏检的可能性增大。针对 IEEE-14 节点系统 27 个状态值模拟二月份第一周的单节点、多节点攻击，攻击状态偏差量 $\boldsymbol{c}$ 为正常状态变量的±10%、±5%，采用准确率、查全率作为阈值选取的调优指标，阈值调整对调优指标的影响如表 5 所示。

Table 5. Detection threshold adjustment based on accuracy and recall
表 5. 基于准确率与查全率的检测阈值调整

检测阈值	准确率	查全率
(0.0010, 0.0025)	0.9859	0.9876
(0.0015, 0.0030)	0.9874	0.9895
(0.0020, 0.0035)	0.9897	0.9910
(0.0025, 0.0040)	0.9903	0.9917
(0.0030, 0.0045)	0.9907	0.9930
(0.0035, 0.0050)	0.9911	0.9874
(0.0040, 0.0055)	0.9922	0.9820

从表中可见，随着检测阈值的增加，攻击检测的准确率缓慢上升，但查全率在电压阈值 τ_1 取 0.003、

相角阈值τ_2取 0.0045 时，呈明显下降趋势，表明随着阈值的增加，攻击漏检的情况显著增多。相比于将正常运行的节点判定为攻击，攻击漏检情况对系统的危害性更强。因此，设定电压阈值τ_1为 0.003、相角阈值τ_2为 0.0045，作为区分节点状态有无受到攻击的评判依据。

4.5. 攻击检测分析

为验证算法性能，将本文检测方法与深度卷积残差神经网络 RestNet50、无迹卡尔曼滤波(unscented kalman filter, UKF)在单节点攻击时的对比见表 6。另外，所提方法在多节点状态攻击检测结果见表 7。

Table 6. False data single node attack detection result
表 6. 虚假数据单节点攻击检测结果

攻击强度	攻击状态	本文准确率	RestNet50准确率	UKF准确率
+5%	θ_2~θ_{14}	0.9796		
	V_1~V_{14}	0.9948		0.9146
+10%	θ_2~θ_{14}	0.9864		(最小值)
	V_1~V_{14}	0.9955	0.9314	
−5%	θ_2~θ_{14}	0.9782		
	V_1~V_{14}	0.9951		0.9872
−10%	θ_2~θ_{14}	0.9852		(最大值)
	V_1~V_{14}	0.9959		

由表 6 可见：

1) 基于 WGAN 状态重构的攻击检测模型针对单节点电压幅值不同强度攻击，其中最小准确率也在 0.99 之上，表明幅值攻击强度对检测模型影响较小。

2) 随着相角攻击强度的衰减，相比于电压幅值，攻击检测准确率略有下降，但整体检测准确率在 0.98 左右波动。其原因为攻击幅度较为接近原始值，且相同缺失比例下电压相角重构精度要低于幅值。

3) 相比 RestNet50 检测方法，本文检测方法凭借可靠的重构效果，整体准确率提升了 6%左右。相比 UKF 检测算法，低强度的相角定位攻击检测准确率也保持在 0.978 之上，表明检测模型良好的稳定性与准确性。

Table 7. False data multi-node attack detection result
表 7. 虚假数据多节点攻击检测结果

攻击状态	−10%准确率	−5%准确率	+5%准确率	+10%准确率
(V_5, V_9)	0.9956	0.9934	0.9946	0.9978
(V_5, θ_9)	0.9821	0.9734	0.9776	0.9830
(θ_5, θ_9)	0.9602	0.9546	0.9523	0.9709
(V_5, V_9, V_{13})	0.9930	0.9926	0.9941	0.9949
(V_5, V_9, θ_{13})	0.9801	0.9702	0.9699	0.9827
(θ_5, θ_9, θ_{13})	0.9584	0.9527	0.9437	0.9563

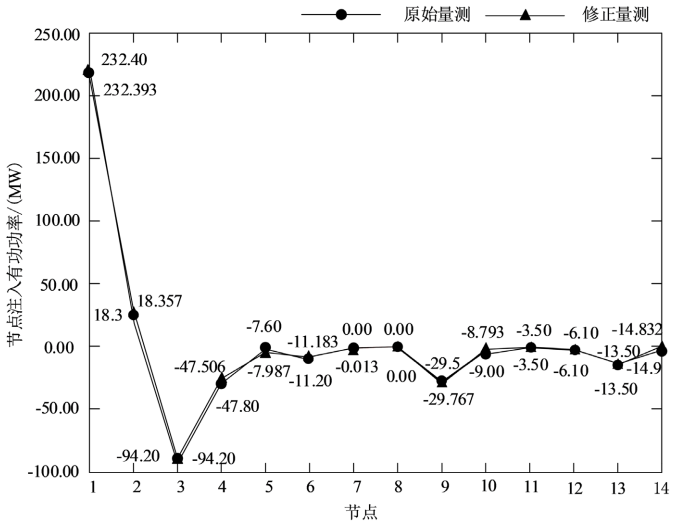
由表 7 可见：

1) 检测模型针对多节点电压幅值不同强度的攻击，检测准确率依旧在 0.99 之上，表明幅值攻击数量对检测模型影响较小。

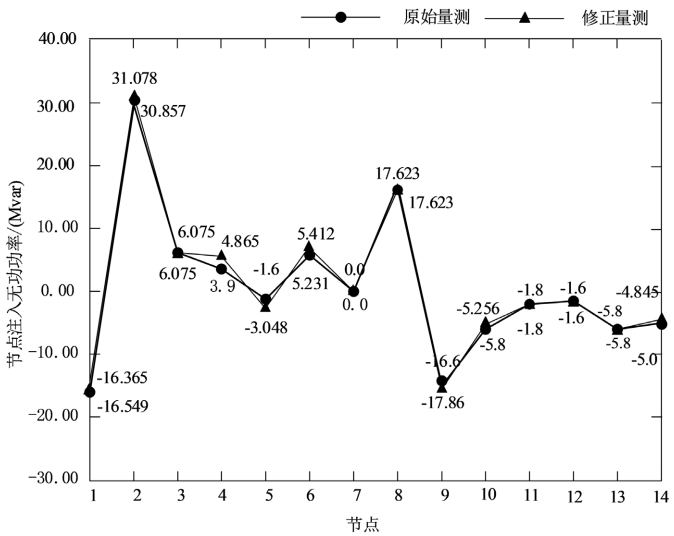
2) 随着相角攻击数量的增加、攻击强度的衰减，检测准确率呈现明显下降趋势，其原因为相角缺失比例的增加，重构误差增大，定位所有相角攻击节点难度进一步增强。

4.6. 修正受损量测

针对多节点攻击，剔除全部虚假数据严重影响电力系统的可观测性。本文利用高精度的重构值替换受损状态，结合网络拓扑结构重新计算修正量测，为电力系统提供接近真实值的伪测量值。以$(\theta_5,\theta_9,\theta_{13})$相角受到状态+5%攻击为例，修正各节点注入有功、节点注入无功如图 6 所示。从图中可得，利用重构状态$(\theta_5',\theta_9',\theta_{13}')$结合网络拓扑结构，所得修正量测与原始量测值吻合程度高，为系统可观测性提供数据支撑。



(a) 节点注入有功功率修正



(b) 节点注入无功功率修正

Figure 6. 3 node voltage phase Angle attack correction result
图 6. 3 节点电压幅值攻击修正结果

5. 结论

为实现虚假数据注入攻击的定位检测与受损量测值修正，提出一种基于状态重构的电力系统虚假数据注入攻击检测模型，并利用 IEEE-14 节点系统和 NYISO 发布的负荷数据进行仿真分析，研究结论如下：

1) WGAN 模型对不同缺失比例状态变量能保证快速重构出接近真实值的状态值，对实现状态攻击检测、修正受损量测提供了稳定的状态参考。

2) 在 IEEE-14 节点系统中，检测模型针对单节点攻击具有较高检测准确率，有效遏制 FDIA，确保电力系统安全稳定运行。

3) 在多节点攻击中，利用高精度的重构状态结合网络拓扑参数提供的修正量测值，确保系统的可观测性。

参考文献

[1] 别朝红, 林超凡, 李更丰, 等. 能源转型下弹性电力系统的发展与展望[J]. 中国电机工程学报, 2020, 40(9): 2735-2745.

[2] 倪明, 颜诘, 柏瑞, 等. 电力系统防恶意信息攻击的思考[J]. 电力系统自动化, 2016, 40(5): 1-4.

[3] Liu, Y., Ning, P. and Reiter, M.K. (2011) False Data Injection Attacks against State Estimation in Electric Power Grids. *ACM Transactions on Information and System Security*, **14**, 13-21. https://doi.org/10.1145/1952982.1952995

[4] 王文钰, 任洲洋, 孙义豪, 等. 基于小波-稀疏自编码器的输电网虚假数据检测方法[J]. 电工电能新技术, 2022, 41(1): 51-59.

[5] Bi, S.Z. and Zhang, Y.J. (2014) Graphical Methods for Defense against False-Data Injection Attacks on Power System State Estimation. *IEEE Transactions on Smart Grid*, **5**, 1216-1227. https://doi.org/10.1109/TSG.2013.2294966

[6] Chao, P., Yang, X. and Wei, L.S. (2020) PMU Placement Protection against Coordinated False Data Injection Attacks in Smart Grid. *IEEE Transactions on Industry Application*, **56**, 4381-4393.

[7] 刘鑫蕊, 吴泽群. 面向智能电网的空间隐蔽型恶性数据注入攻击在线防御研究[J]. 中国电机工程学报, 2020, 40(8): 2546-2559.

[8] Gu, C.J., Panida, J. and Mehul, M. (2015) Detecting False Data Injection Attacks in AC State Estimation. *IEEE Transactions on Smart Grid*, **6**, 2476-2483. https://doi.org/10.1109/TSG.2015.2388545

[9] Wang, D.F., Wang, X.J., Zhang, Y., *et al*. (2019) Detection of Power Grid Disturbances and Cyber-Attacks Based on Machine Learning. *Journal of Information Security and Applications*, **46**, 42-52. https://doi.org/10.1016/j.jisa.2019.02.008

[10] Xue, D.B. and Jing, X.R. (2019) Detection of False Data Injection Attacks in Smart Grid Utilizing ELM-Based OCON Framework. *IEEE Access*, **7**, 31762-31733. https://doi.org/10.1109/ACCESS.2019.2902910

[11] 朱杰, 张葛祥. 基于历史数据库的电力系统状态估计欺诈性数据防御[J]. 电网技术, 2016, 40(6): 1772-1778.

[12] 杨怡, 王勇. 基于 AUKF 的分布式电源系统虚假数据攻击检测方法[J]. 电工电能新技术, 2021, 40(12): 48-55.

[13] 杨玉莲, 齐林海, 王红, 等. 基于生成对抗网络和双重语义感知的配电网量测数据缺失重构[J]. 电力系统自动化, 2020, 44(18): 46-54.

[14] 王守相, 陈海文, 潘志新, 等. 采用改进生成式对抗网络的电力系统量测缺失数据重建方法[J]. 中国电机工程学报, 2019, 39(1): 56-64.

[15] 郑文迪, 聂建雄, 邵振国, 等. 智能配电网状态估计研究现状和展望[J]. 电力系统及其自动化学报, 2021, 33(4): 8-16.

[16] 王电钢, 黄林, 刘捷, 等. 考虑负荷虚假数据注入攻击的电力信息物理系统防御策略[J]. 电力系统保护与控制, 2019, 47(1): 28-34.

[17] 赵丽莉, 刘忠喜, 孙国强, 等. 基于非线性状态估计的虚假数据注入攻击代价分析[J]. 电力系统保护与控制, 2019, 47(19): 38-45.

[18] Ledig, C. and Theis, L. (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *Proceedings of* 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 105-114. https://doi.org/10.1109/CVPR.2017.19

[19] Qian, S., Liu, H., Liu, C., *et al*. (2018) Adaptive Activation Functions in Convolutional Neural Networks. *Neurocom-*

puting, **272**, 204-212. https://doi.org/10.1016/j.neucom.2017.06.070

[20] Load Data: Market and Operational Data (NYISO). http://www.energyonline.com/Data/GenericData.aspx?DataId=13&NYISO___Hourly_Actual_Load