基于 PanSec-GNN 框架的新型配电系统终端身份欺骗检测研究

胡兴元

中检集团天帷信息技术（安徽）股份有限公司，安徽合肥，231200；

摘要：配电级 AMI 的大规模部署使智能电表成为身份欺骗的新攻击面。传统加密与阈值检测难以应对凭证泄露及间歇伪装引发的拓扑冲突。针对上述问题，提出一种 PanSec-GNN 框架：以 AMI 通信-电气耦合拓扑构建动态图，将欺骗检测转化为节点级异常分类。框架递进式引入 GCN 捕获邻域一致性、GAT 学习边权重聚焦可疑交互，并以 GRU-残差 STGNN 建模时序演化。IEEE-34 节点数据集上的 7 天 50 次多策略攻击实验表明，PanSec-GNN 取得 96.1% 准确率、94.8%F1，显著优于现有基线；注意力可视化可直接定位身份冲突链路，为运维人员提供可解释告警。

关键词：智能电网安全；身份伪造攻击；图神经网络；注意力机制；时空建模

DOI：10.69979/3060-8767.25.09.004

引言

现代智能电网依托先进的计量基础设施（AMI）实时监测并控制电力分配，智能电表一般安装在居民区和变电站，依靠密集的传感器网络，给公用事业提供商传递用电数据及控制信号，尽管此互联架构增进了运营效率，但也增多了网络威胁出现的几率[1]。身份顶替是一种极其隐蔽的威胁，恶意设备借助合法智能电表的凭证冒充其电表身份[2]，一旦系统接收了冒名的身份，攻击者即可输入伪造的测量值或者控制指令，也许会引发电网运行中断或操纵计费系统，传统的安全手段可阻止无许可的访问[3]；复杂的欺骗攻击说不定能绕过这些防御手段[4]，尤其是当内部凭证被泄露，或者传统协议（例如 Modbus[5]或 DNP3[6]）没有身份验证机制的时候。此挑战凸显出入侵检测系统（IDS）的不可或缺性，该系统借助分析电网里设备的行为模式来识别反常情况，以此查找身份欺骗的迹象，图神经网络（GNN）具备很强的关系数据建模实力，可应用在智能电网环境下的安全风险分析[7]，智能电网的物理跟通信基础设施自然演化出一个图结构，其中节点指代设备，边代表了通信与电气方面的连接。身份伪造攻击大多引入细微的拓扑不一致或异常交互模式，但不改变单个个体的数据值，纵使传统点对点方法宣告失效，GNNs 依旧能在电网拓扑的上下文中剖析电表数据并识别此类异常，近期网络物理安全领域相关研究表明[8]，引入图结构可明显提升对虚假数据注入等威胁的检测本领，目前的研究大多聚焦于广域攻击，忽视了图的依赖关系，本研究重点聚焦配电网层面的身份欺骗问题，又提出一种基于 GNN 的检测器，目标是捕捉此类攻击背后渐趋复杂的模式。

1 方法

本研究提出的 PanSec-GNN 网络是在智能配电网络的控制架构中运行，其中智能电表通过 AMI 通信网络定期传输读数和 ID。在身份伪造攻击中，攻击者利用窃取的电表凭证注入伪造数据，这些数据可能来自不同位置或以不规则方式注入。为检测此类威胁，构建一个通信图，其中节点代表设备，边代表网络或电气连接，如图 1 所示，其中 BR 表示断路由器，R 表示继电器，G 表示发电机。每个节点关联有功率读数、时间戳和编码 ID 等特征。IDS 实时构建该图并将其输入基于 GNN 的模型进行节点级分类。通过捕获网络中的拓扑不一致性和身份冲突，PanSec-GNN 能够有效识别在孤立状态下看似正常但在特定上下文中表现异常的伪造设备。

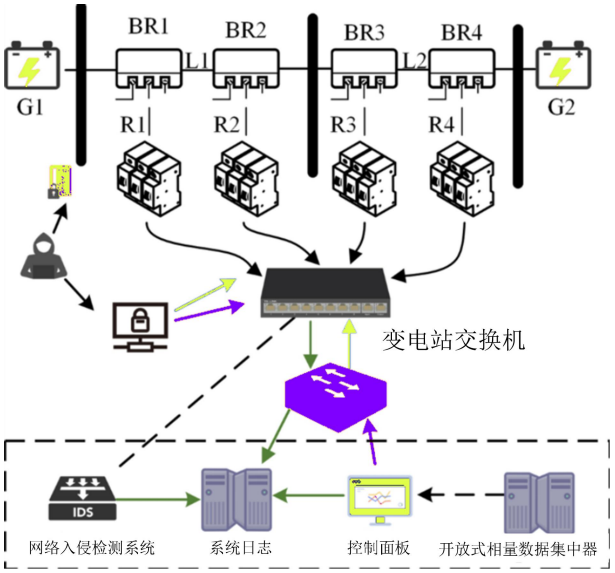

图 1. PanSec-GNN 框架

1.1 GCN 检测器

本研究通过电力分配网络的通信拓扑结构，构建了

一个基线GCN模型来描述智能电表之间的空间依赖关系。图中的每个节点代表一个智能电表，并关联一个特征向量 $\mathbf{x}_i$，其中包含最近的电能读数、传输时间戳及身份信息。GCN 通过邻域聚合机制学习节点表示。具体而言，对于图 $G=(V,E)$ 有邻接矩阵 A 和度矩阵 D，第 l 层的 GCN 传播规则定义为：

$$\mathbf{H}^{(l+1)}=\sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \quad (1)$$

其中，$\mathbf{H}^{(l)}$ 是第 l 层的输入特征矩阵，$\mathbf{W}^{(l)}$ 是可训练的权重矩阵，是非线性激活函数，例如 ReLU。经过两层图卷积后，得到节点嵌入向量 $\mathbf{H}^{(2)}$ 然后将这些嵌入向量输入到 sigmoid 分类器中，以生成每个节点的欺骗概率分数：

$$\hat{y}_i=\text{sigmoid}\left(\mathbf{w}\ \ \mathbf{h}_i^{(2)}\right) \quad (2)$$

这里，$\mathbf{h}_i^{(2)}$ 表示节点 i 的最终表示，，$\mathbf{w}$ 是可训练的分类向量。接下来，模型使用二分类交叉熵损失在标注节点上进行训练：

$$\mathrm{L}_{\mathrm{BCE}}=-\sum_{i\in \mathrm{V}_{\mathrm{train}}}\left[y_i\log\hat{y}_i+(1-y_i)\log(1-\hat{y}_i)\right] \quad (3)$$

GCN 允许每个节点在其网络邻居的上下文中推断其行为的一致性，从而为检测身份欺骗攻击建立了拓扑感知的基础。

1.2 GAT 检测器

为了提升模型区分智能电表间关键关系的能力，扩展 GCN，引入了注意力机制，并构建了基于 GAT 的检测器。与 GCN 不同，GCN 通过固定的归一化邻接矩阵聚合邻居信息，而 GAT 允许模型为每条边学习自适应权重，从而在更新节点表示时更强调信息丰富的邻居节点。具体而言，给定节点 i 及其邻居节点 $j\in N(i)$，非归一化注意力系数 e_{ij} 的计算公式如下：

$$e_{ij}=\text{LeakyReLU}\left(\mathbf{a}\ \ [\mathbf{W}\mathbf{h}_i\|\mathbf{W}\mathbf{h}_j]\right) \quad (4)$$

其中 h_i 和代表结点 i 和 j 的特征，并且 a 是一个可学习的权重向量。然后，使用 softmax 函数对所有邻居的注意力权重进行归一化：

$$\alpha_{ij}=\frac{\exp(e_{ij})}{\sum_{k\in \mathrm{N}\ (i)}\exp(e_{ik})} \quad (5)$$

这些归一化权重 α_{ij} 决定了邻居节点 j 对节点更新后的表示的重要性，节点 i 的新表示通过加权求和计算得出：

$$\mathbf{h}_i'=\sigma\left(\sum_{j\in \mathrm{N}\ (i)}\alpha_{ij}\mathbf{W}\mathbf{h}_j\right) \quad (6)$$

图 2 展示了基于 GAT 的检测器的架构。每个智能电表（节点）利用学习到的边权重 α_{ij} 来关注其邻近节点，从而使模型能够对可疑通信链路赋予更高的优先级。

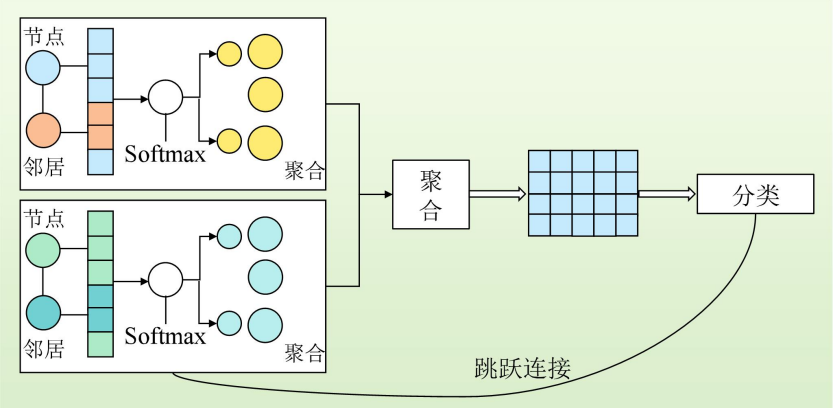

图 2. 基于 GAT 的检测器

1.3 STGNN 检查器

为了进一步提升智能电网中身份欺骗的检测能力，引入一种基于 GAT 的时序建模机制，并构建了 STGNN[13]。考虑到身份欺骗通常并非瞬间发生，而是可能以间歇性伪装、周期性注入或在一定时间内逐步调整数据值等形式出现，以逃避传统检测方法，改进后的 STGNN 设计能够有效捕捉此类时间连续特征 。该模型以一系列图 $G^{(t-T+1)},\ldots,G^{(t)}$ 作为输入。首先，它使用 GAT 在每个时间步 τ 取每个节点的结构特征。然后，它将时间序列表示 $\mathbf{z}_i^{t-T+1},\ldots,\mathbf{z}_i^t$ 输入到门控循环单元（GRUs）中，以建模其演化过程，最终获得每个节点的时敏表示 $\mathbf{s}_i$。

在模型的训练阶段，STGNN 凭借监督学习开展优化，各个时间步中的每个节点被标记为正常或欺诈，进而实现二分类这一任务，为了增进模型对极端异常值的鲁棒

程度，在输入跟输出之间引入残差连接，让模型在必要情况下可直接依靠当前读取的信息，从而显著增强其对严重伪造行为的响应本领。STGNN 在空间关系建模以及时间序列动态感知间实现平衡，大幅增强了识别隐蔽身份伪造攻击的水平，且有效减少了异常用户行为等非恶意场景里的误报数量，这种把空间、时间上下文联合起来的建模途径，能适配智能电力分配网络中不断转变且高度相关的终端场景。

2 实验

2.1 实验细节和数据集

所选用的模型借助 Python 和 PyTorch 进行构建，采用 Adam 做优化，且运用早期停止和 L2 正则化训练以杜绝过拟合现象，为了增进泛化能力，在训练里融入了不同的身份伪造场景，包含单点攻击、多点攻击以及时间部分相关攻击，STGNN 采用的时间窗口是 T 取值 10，拥有 64 维 GNN 隐藏状态以及 4 个 GAT 头节点，这些参数经由网格搜索得以确定。

依靠 IEEE34 节点拓扑结构对一个智能电网做了模拟，于 10 个社区里部署了 80 个智能电表，还采用 PecanStreet 数据集中的实际用电轨迹来模拟正常的用电表现，在 10%的时间步里实施欺骗攻击，引入具有重复 ID、运用多种欺骗策略的恶意节点：隐蔽类。每一次攻击都会引起电表拓扑结构或时间序列的变动，让检测面临难题，真实标签对欺骗节点与被入侵的合法电表做了标记，在长达 7 天的模拟阶段，实施了约 50 次攻击行动，影响到了大概 30 个电表，其中包含部分合法 ID 的修改，以此测试抵抗误报能力。

2.2 对比结果

通过对比所提出的 PanSec-GNN 与多种基线方法，包括传统机器学习模型和近期基于图的方法，如表 1 和图 3 所示。实验结果表明，基于 STGNN 的检测器具有显著优势，尤其在更具挑战性的节点级欺骗检测任务中表现突出。无监督方法如孤立森林（IF）无法捕捉复杂的拓扑依赖关系和时序相关性，导致准确率低且假阳性率高。监督基线方法如支持向量机（SVM）和长短期记忆网络（LSTM）通过融入标注信息或时序结构略有改善，但仍缺乏对仪器交互的感知能力。先进的图基基线方法如 GCN-AE 和 GraphKAN 通过利用结构信息提升性能。然而，所提 STGNN 进一步整合了电表行为的空间和时间维度。具体而言，PanSec-GNN 实现了节点级检测准确率 96.1%和 F1 分数 94.8%，超越所有基线方法。

表 1. 与基线方法在欺骗检测中的对比结果(%)

方法	准确率	F1 分数	AUC
IsolationForest(IF)	78.2	71.4	80.5
SVM	84.2	76.3	86.1
LSTM	87.5	80.5	88.9
DNN	82.3	75.1	84.2
GCN-AE	90.1	86.7	91.5
GraphKAN	92.2	89.0	93.3
Ours	96.1	94.8	97.5

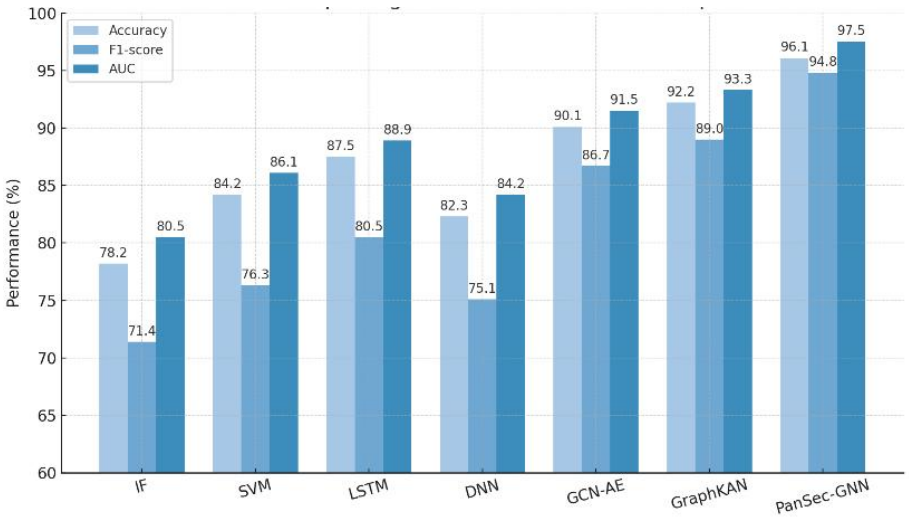

图 3. 不同方法的欺骗检测性能对比

2.3 消融实验

为了评估 PanSec-GNN 中各组件的贡献，通过有针对性地移除或修改关键模块进行了消融实验。如图 4 和表 2 所示，测试该模型的三个简化变体：仅 GCN 版本移除了注意力机制和时间建模，仅使用标准图卷积在单个快照上进行处理。移除 GAT 的变体保留了空间建模但移除了时间组件（如 GRU），能够独立处理每个时间步。Temporal-GRU-removed 版本保留了图结构和注意力机制，但移除了残差连接并简化了时间学习组件。在所有配置下，所提模型以 F1 分数 94.8%取得了最佳性能。移除时间 GRU 导致性能适度下降至 93.1%，表明捕获时间动态可提升检测效果，尤其对渐进或间歇性欺骗更有效。完全禁用时间建模（不使用 GAT）导致准确率进一步下降至 91.7%，证实了时空集成的重要性。性能下降最显著的设置是仅使用 GCN 的场景，F1 分数降至 89.0%，这突显了图拓扑结构和注意力机制在检测身份欺骗中的重要性。这些结果表明，空间关系、基于注意力的邻域权重以及时序行为建模共同作用，为 PanSec-GNN 框架提供了鲁棒且准确的检测能力。

表 2. PanSec-GNN 中各个组件的消融实验 (%)

方法	准确率	F1 分数	AUC
GCNonly	91.2	89.0	92.6
w/oGAT(notemporalmodeling)	93.8	91.7	95.2
w/oTemporalGRU(noresidual)	95.0	93.1	96.3
Ours	96.1	94.8	97.5

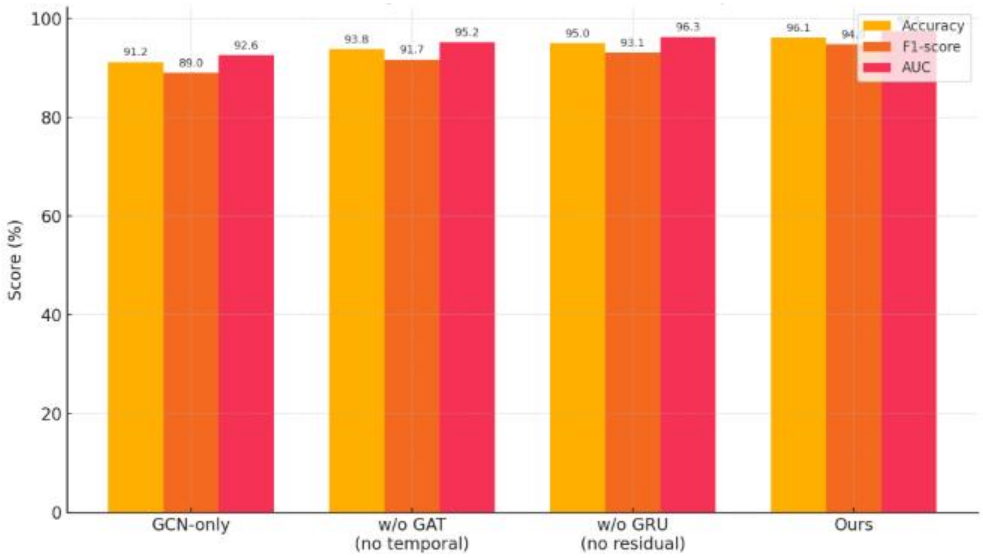

图 4. PanSec-GNN 组件的消融研究结果

2.4 实验结果可视化

基于注意力机制的图神经网络（GNNs）的一个关键优势在于，它们能够解释模型在标签攻击过程中“关注”的对象。图 5 展示了测试集中的一个典型攻击事件，即一种隐蔽的类型 I 欺骗攻击，攻击者引入了一个幽灵电表来复制 ID 为 37 的电表读数。在此子图中，节点 0 是攻击者的恶意电表，节点 9（绿色）是具有相同 ID 的合法电表。模型正确地将节点 0 标记为恶意电表（红色）。节点 0 与节点 9 之间的边（通过网络连接）具有高注意力权重（橙色显示），表明模型已学会关注这些节点之间的交互（或冲突）。虚线灰色边（正常权重）连接其他对决策影响较小的电表。这符合直觉：间接地，主要线索是节点 0 和 9 之间的身份冲突。事实上，模型在最终的 GAT 层中，将连接这两个节点的消息权重分配了近 0.9。从操作员的角度来看，系统可以大致说明：“变压器 X 上的电表 37 被标记，因为它也出现在变压器 Y 上”，从而引导工程师识别问题的本质。

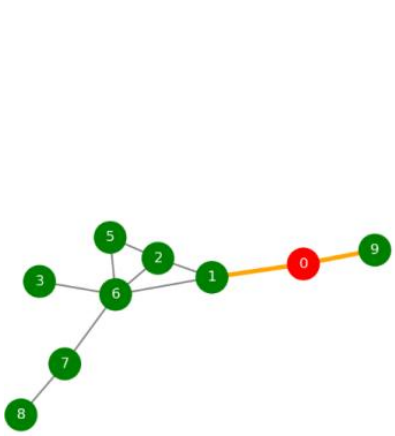

图 5. 典型攻击事件案例

3 结论

本研究提出了一种以图神经网络（GNN）为基础的入侵检测框架 PanSec-GNN，意图检测并消除智能电表网络里的身份冒充攻击，把冒充检测问题建模成图上的节点分类任务，该框架全面利用了智能电网里通信和物理拓扑结构形成的关联性和时序不一致性。依靠逐步开展的模型设计，从 GCN 过渡到 GAT，最终到结合 GRU 的时空图神经架构，实验体现了检测性能的不断增强，最终模型明显胜过传统基线方法，含有孤立森林、SVM 以及近期基于图的检测器具，该模型凭借注意力机制的可视化达成了可解释性，为电网运营商提供了针对身份报告冲突等可疑行为的实用觉察。

参考文献

[1]ZibaeiradA,KoleiniF,BiS,etal.Acomprehensive surveyonthesecurityofsmartgrid:Challenges,mitigations,andfutureresearchopportunities[J].arXivpreprintarXiv:2407.07966,2024.

[2]刘生源,游书堂,尹浩等.电力系统网络安全中的无模型数据认证[J].智能电网汇刊,2020,11(5):4565-4568.

[3]崔宇,白峰,刘勇,等.智能电网中同步相量数据对抗欺骗攻击的时空特征分析[J].电气与电子工程师学会智能电网汇刊,2019,10(5):5807-5818.

[4]EastS,ButtsJ,PapaM,etal.ATaxonomyofAttacksontheDNP3Protocol[C]//InternationalConferenceonCriticalInfrastructureProtection.Berlin,Heidelberg:SpringerBerlinHeidelberg,2009:67-81.

[5]Radoglou-GrammatikisP,SiniosoglouI,LiatifisT,etal.Implementationanddetectionofmodbuscyberattacks[C]//20209thInternationalConferenceonModernCircuitsandSystemsTechnologies(MOCAST).IEEE,2020:1-4.

[6]AbineshR,VGY,TJS,etal.Deepgraphconvolutionneuralnetworkbasedintrusiondetectionsystemtowardsearlydetectionofmaliciousattacks[C]//20248thInternationalConferenceonI-SMAC(IoTinSocial,Mobile,AnalyticsandCloud)(I-SMAC).IEEE,2024:549-554.

[7]TakiddinA,AtatR,IsmailM,etal.Generalizedgraphneuralnetwork-baseddetectionoffalsedatainjectionattacksinsmartgrids[J].IEEETransactionsonEmergingTopicsinComputationalIntelligence,2023,7(3):618-630.

[8]SweetenJ,TakiddinA,IsmailM,etal.Cyber-physicalgnn-basedintrusiondetectioninsmartpowergrids[C]//2023IEEEinternationalconferenceoncommunications,control,andcomputingtechnologiesforsmartgrids(SmartGridComm).IEEE,2023:1-6.