FedDiSC：一种面向电力系统扰动和网络攻击判别的计算高效联邦学习框架⋆

穆罕默德·阿克巴尔 胡斯努[a]，*，2， 阿德南 安瓦尔[a]， 哈夫图·塔瑟乌·雷达[a]， 纳赛尔 侯赛因扎德[b]， 沙玛·纳兹 伊斯兰[b]， 阿卜敦·纳瑟尔马哈茂德[c] 和 罗宾 多斯[a]，*

[a]网络安全研究与创新中心（*CSRI*），信息技术学院，迪肯大学，澳大利亚维多利亚州吉朗*3216*[b]智能电力与能源研究中心（*CSPER*），工程学院，迪肯大学，澳大利亚维多利亚州吉朗*3216*[c]计算机科学与信息技术系，拉筹伯大学，澳大利亚维多利亚州邦多拉*3086*

文章信息

关键词: 联邦学习 异常检测 网络攻击 物联网（IoT） 智能电网

摘要

随着人们日益关注智能电网系统的安全与隐私，针对电网关键组件（如状态估计）的网络攻击已被证明是网络安全领域的首要问题之一，近年来备受关注。然而，智能电网中的网络攻击检测如今面临新的挑战，包括隐私保护，以及存在具有战略性的数据所有者的分散式电力区域。为解决这些技术瓶颈，本文提出一种新颖的、基于**联邦**学习的隐私保护且通信高效的攻击检测框架，称为*FedDiSC*，可区分电力**系统**扰动与**网**络**攻击**。具体而言，我们首先提出一种联邦学习方法，使分散式电网区域的监控控制与数据采集子系统能够协同训练攻击检测模型，而无需共享敏感的电力相关数据。其次，我们提出一种基于表征学习的深度自编码器网络，用于准确检测电力系统异常和网络安全异常。最后，为使所提出的框架适应智能电网中现实网络攻击检测对时效性的要求，我们采用一种称为 DP-SIGNSGD 的梯度隐私保护量化方案，以提高其通信效率。针对公开可用的工业控制系统数据集对所提出框架进行的大量仿真表明，该框架能够在保护电网敏感相关信息隐私的同时，实现更高的检测准确率。此外，我们发现，与不采用梯度量化的传统联邦学习方法相比，所采用的梯度量化方案可将通信效率提高 40%，表明该方案适用于现实场景。

arXiv:2304.03640v1 [cs.CR] 7 Apr 2023

1. 引言

智能电网（SG）范式中物联网（IoT）的广泛应用，使边缘设备生成的数据量以前所未有的速度增长。这些数据对电力系统的诸多关键运行至关重要，包括需求预测、电网优化、电力系统监测与控制等。因此，能源电网系统的整体管理能够更加动态、高效地开展。然而，电力系统基础设施异构性、多样性和复杂性的增加，对其完整性构成了严峻挑战。确保数据完整性是现代电力系统的核心任务之一 [1]。这是因为，现代网络物理电网组件无处不在，加之信息技术不断进步，使得如今能够在边缘侧进行数据处理和决策 [2]，而传统方式则依赖云存储型控制中心。

数字技术的进步推动了能源关键基础设施的重大改善，从而缓解了现代世界面临的最关键问题——如何满足永无止境的能源需求。尽管智能电网具有显著的优势，但仍面临诸多技术挑战。虽然智能传感设备在智能电网范式中发挥着关键作用，但网络攻击可能严重影响系统的完整性和韧性 [3]。智能电网的许多信息物理组件都可能成为网络攻击的潜在目标 [4]。其中一种关键的网络攻击是 虚假数据注入 （FDI），这是一种关键的网络攻击，恶意攻击者通过篡改原始传感器测量值或状态估计结果发起严重攻击 [5]。此外，攻击者不断改进攻击构造方法，将遭到破坏的测量值伪装成自然扰动/故障，使得在恶意事件与非恶意事件之间进行检测变得极其困难，甚至令坏数据检测方法无法实现 [6]。因此，设计有效的防御措施以保护现代电力系统免受此类网络威胁至关重要。

最近，已有多项研究致力于开发针对智能电网上网络攻击的防御措施。以往的工作主要集中于利用机器学习算法对智能电网系统进行网络攻击检测。例如，[7]中的工作利用基于监督学习的分布式支持向量机，结合交替方向乘子法，并取得了令人满意的-

⋆本研究是由迪肯大学科学、工程与建筑环境学院（SEBE）通过“Mini ARC Analogue计划（MAAP）”资助的研究项目成果。

mahusnoo@deakin.edu.au (M.A. Husnoo);adnan.anwar@deakin.edu.au (A. Anwar);haftu.reda@deakin.edu.au (H.T. Reda);nasser.hosseinzadeh@deakin.edu.au (N. Hosseinzadeh);shama.i@deakin.edu.au (S.N. Islam); A.Mahmood@latrobe.edu.au (A.N. Mahmood);robin.doss@deakin.edu.au (R. Doss) ORCID（学术档案）编号：0000-0001-7908-8807 (M.A. Husnoo); 0000-0003-3916-1381(A. Anwar);0000-0001-7769-3384 (H.T. Reda); 0000-0002-2354-7960 (S.N. Islam); 0000-0001-7769-3384 (A.N. Mahmood)

检测性能更出色且计算复杂度更低，在 IEEE 标准测试系统上进行测试时表现尤为明显。Anwar、Mahmood 和 Shah 的另一篇文献 [8] 声称，他们基于主成分分析和序列最小优化（SMO）的支持向量方法所提出的解决方案，在工业控制系统数据集上进行测试时，性能优于广泛使用的机器学习算法。最近，鉴于深度学习方法相较于传统机器学习算法具有更高的准确率，研究人员开始将其用于攻击检测。例如，在[9], 该研究中，作者开发了一种无监督深度信念网络
基于检测的方法在 IEEE标准测试系统上进行测试时取得了令人满意的结果。另一方面，[10]提出了一种先进的深度集成学习算法，以改进住宅需求响应场景中的虚假数据注入检测。然而，众多既有文献中提出的集中式网络攻击检测技术存在以下问题：1）存储能力有限，2）通信瓶颈，以及最重要的 3）隐私问题。研究人员往往忽视这样一个事实：电力系统数据可能包含敏感信息，而将其传输至集中式控制中心最终可能导致隐私泄露 [11]。

为应对上述挑战，研究人员提出了一种有前景的、基于隐私保护分布式学习并利用联邦学习（FL）的网络攻击检测方法。这种新型机器学习范式使边缘设备能够利用本地训练样本，在边缘侧设备上训练网络攻击检测深度学习神经网络模型，同时无需共享其本地数据集，从而保护隐私。例如，[12] 提出了一个用于检测太阳能光伏直流/直流和直流/交流变流器中虚假数据注入（FDI）攻击的联邦学习框架。同样，[13] 开发了一种基于激励机制的联邦攻击检测方法，用于电网状态估计，据称其性能优于现有研究。[14] 的作者实现了一种基于联邦Transformer的虚假数据注入（FDI）检测系统，并在IEEE标准测试系统上进行了广泛验证，结果表明其所提机制具有有效性。然而，目前专注于利用联邦学习检测电力控制系统中隐蔽网络威胁的研究仍然十分有限。

在这项工作中，我们提出采用联邦式网络攻击检测机制，以区分恶意与非恶意的电力系统扰动，从而减轻电力系统操作员判断某一电力扰动是否为有意行为的负担，并提供可靠的决策支持。我们建立了分布式电网区域，使SCADA子系统能够在本地检测网络威胁，而无需将数据上传至控制中心，从而解决集中式方法面临的存储能力、通信可靠性和隐私问题。具体而言，本文的主要贡献总结如下：

1. 我们提出了一种基于联邦学习的轻量级、保护隐私的设备端协同深度网络攻击检测框架，适用于电力系统网络，称为 *Fed-DiSC.*

2. 我们提出了一种基于表征学习的深度去噪自编码器（DAE）网络，用于准确检测电力系统异常和网络安全异常。

3. 我们采用 DP-SIGNSGD [15], 这一隐私保护的梯度量化方案，以提高所提出框架的计算效率。该方案通过减少客户端节点与中央协调器之间交换的参数数量，降低了通信开销。

4. 我们在公开可用的工业控制系统数据集上进行了广泛实验，以验证我们提出的框架确实能够在保持较低通信开销的同时，以高准确率检测异常。我们将所提出的框架与第4.2 节中讨论的多种基于联邦学习的深度学习模型和集中式机器学习模型进行了比较，结果发现，我们的攻击检测框架达到了94.8%的更高准确率，并将训练时间缩短了40%。

本文其余部分的结构如下：第2节介绍网络攻击检测和联邦学习领域已有的前沿研究成果。第*3*节介绍所提出的FedDiSC方法，包括系统模型、检测模块等。第4节介绍用于验证所提方法在公开数据集上有效性的仿真场景。最后，我们在第5节总结全文。

2. 相关工作

下文将网络攻击检测和联邦学习领域当前的前沿研究进行总结，并将以往相关文献分为两大类：数据驱动的网络攻击检测，以及与智能电网相关的联邦学习。

2.1. 数据驱动的网络攻击检测

尽管网络攻击检测有多种方法和研究方向，但基于模型的检测算法凭借其可扩展性、实时检测能力以及不依赖系统参数的特点，在文献中受到了广泛关注 [16]。例如，文献 [17] 是最早尝试利用基于先验信息的高斯过程回归建模技术，开发预警系统以减轻网络攻击对智能电网影响的研究之一。另一项研究 [4] 在工业控制数据集上训练了随机森林、朴素贝叶斯等常见机器学习模型，并取得了较高的检测性能。该研究还进一步探索了文献 [18]中随机森林、J48、JRip等不同基准分类器的应用。Hao、Piechocki、Kaleshi、Chin和Fan[19] 提出采用鲁棒主成分分析（RPCA）方法检测随机和定向虚假数据注入攻击。同样，Yang、Wang、

Zhou、Ruan和Liu [20]采用了四种传统机器学习模型，即单类支持向量机、鲁棒协方差、孤立森林和局部离群因子方法来检测网络攻击，并在 IEEE 14 母线系统上进行仿真后取得了较高的分类性能。Kumar、Saxena和Choi开展的另一项近期研究[21] 通过在电力系统数据集上训练 AdaBoost、梯度提升、近邻等多种常见机器学习模型，解决了网络攻击检测问题，检测准确率达到 70% 至 92%。

近年来，随着深度学习应用呈指数级增长，研究人员正积极采用更新的深度学习算法，以进一步推进和改进智能电网中的网络威胁检测。例如，文献[22] 开发了一种基于多层感知器的方法，用于增强智能电网系统抵御虚假数据注入攻击的韧性。Niu、Li、Sun和Tomsovic 的另一项研究 [23]开发了一种用于智能电网异常检测的CNN-LSTM 网络，其性能优于传统的坏数据检测器。类似地，文献 [24] 中的作者采用最先进的 CNN 作为多标签分类器，以防御潜在的受篡改测量数据威胁。Wang、Zhang、Ma 和Jin [25] 提出将卡尔曼滤波器与循环神经网络相结合，该方法显著提高了针对智能电网网络攻击的检测率。最近，Bitirgen 和 Filik [26] 提出了一种结合粒子群优化的 CNN-LSTM 集成模型，与现有研究相比，该模型实现了较高的网络攻击检测准确率。然而，当不同的系统运营商管理相互独立的电网区域时，会产生隐私问题，而这在电力系统网络中是非常常见的场景。此外，上述大部分文献并未解决在利用此类敏感的电力相关数据进行网络威胁检测时的用户隐私问题。不同于现有研究，我们提出的隐私保护
结合联邦学习的梯度量化方案，通过支持在设备上以计算高效的方式协同训练攻击检测模型，解决了这一隐私问题，从而区分电力系统扰动与网络攻击。

2.2. 联邦学习在智能电网应用中的应用

联邦学习是一种根本性的新型机器学习范式，能够在设备端协作进行去中心化模型训练，而无需将原始数据物理迁移至集中式服务器 [27]。自诞生以来，联邦学习已应用于多个对隐私保护要求极高的智能电网应用中。Taik和 Cherkaoui [28] 开创性地将联邦学习应用于智能电网领域，设计了一种无需共享数据的隐私保护型需求负荷预测框架。他们在一个真实世界的德克萨斯州负荷消耗数据集上训练了一个 LSTM模型，并取得了足够的预测性能，从而证明了联邦学习在智能电网领域的有效性。 [29] 中的另一项研究利用联邦学习进行分布式能源资源预-测，并验证了其较高的预测性能，使用-ing 使用 GridLAB-D 仿真和 Pecan Street数据集。类似地，Gholizadeh 和 Musilek [30] 以及 Yang、Wang、Zhao 和 Wu [31] 在采用联邦学习（FL）进行需求预测后，取得了令人满意的负荷预测性能。由于其在需求预测应用中的突出表现和优异性能，联邦学习（FL）很快在智能电网领域的其他方面获得了关注。例如，文献 [32] 的作者设计了一种用于能源窃取检测的联邦时序卷积网络。他们声称，在使用真实能源消耗数据集进行广泛的数据驱动实验后，其联邦框架能够以较小的计算开销实现较高的检测准确率。在类似的研究中，Ashraf、Waqas、Abbas、Baker、Abbas 和 Alasmary[33]在真实世界数据集上采用联邦传统机器学习算法后，在能源窃取检测方面取得了相当的性能。最近，与联邦网络攻击检测相关的研究还很少。文献 [12] 提出了一种联邦学习（FL）框架，用于检测太阳能光伏 DC/DC和DC/AC变换器上的虚假数据注入（FDI）攻击。类似地， [13]开发了一种基于激励机制的联邦虚假数据注入攻击检测方法，用于电网状态估计，并声称其性能优于现有研究。文献 [14]实现了一种基于联邦 Transformer的虚假数据注入（FDI）检测系统，并在 IEEE 标准测试系统上进行了广泛验证，结果表明其所提出机制的有效性。相比之下，本文的重点是开发一种面向电力控制系统的全电网范围异常检测框架，该框架能够区分非恶意电力系统扰动和恶意虚假数据注入（FDI）威胁，同时具备较高的计算效率。

3. 所提出的方法

本节将首先介绍针对该场景所考虑的系统模型，随后介绍我们提出的基于计算高效联邦学习的隐私保护入侵检测框架。

3.1. 系统模型与假设

在本文中，我们考虑一种通用的按区域分布式电网系统，该系统包含一个控制中心和 K 基于边缘的监控与数据采集（SCADA）子系统，这些子系统协同训练虚假数据注入攻击检测模型，如图1所示。基于边缘的SCADA子系统利用各自的本地数据集在本地训练上述共享模型，这些本地数据集由各自电网区域中不同IED传感节点采集。完成本地训练后，各子系统将模型更新共享至控制中心，由控制中心采用多数投票聚合算法对模型更新进行平均，从而获得更新后的全局模型。随后，更新后的模型梯度被共享回边缘SCADA子系统，并用于虚假数据注入攻击检测。我们所提出的系统模型包含以下三个关键参与方，现简要介绍如下：

1) 控制中心：控制中心是监控和协调电网运行的中央设施。通常，

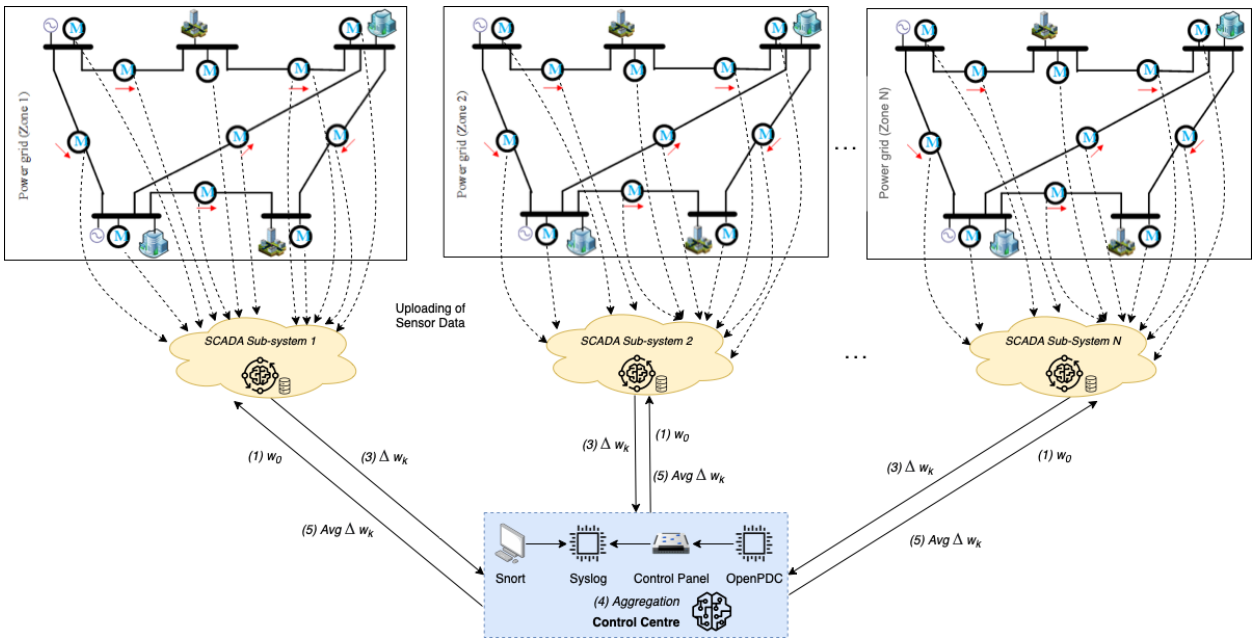


图 1：基于联邦学习的入侵检测框架概览。我们提出的框架的直观工作流程如下：1）电网区域内的边缘智能电子设备将数据安全地上传至 SCADA 子系统。2）控制中心向所有 SCADA 子系统分发统一的模型和参数初始化设置。3）各 SCADA 子系统使用各自的本地数据集训练本地模型。4）各 SCADA 子系统将压缩梯度更新发送至控制中心。5）控制中心聚合上传的压缩梯度，以获得最新模型。6）控制中心将模型更新发送回各 SCADA 子系统。重复步骤 1–6，直至全局模型达到最优收敛。

控制中心拥有云服务器和计算资源，能够提供充足的计算能力，充当联邦学习架构中的中央协调器。控制中心主要负责初始化全局模型、向边缘侧的SCADA子系统广播模型参数，最后汇总边缘侧客户端共享的模型梯度。我们假设控制中心是诚实但好奇的参与方，这意味着控制中心会试图尽可能多地获取信息，但不会采取对抗行为。

2）*SCADA* 子系统：SCADA 子系统是具备先进数据采集能力的边缘客户端，在电网系统中负责采集和监控各传感器设备（即 IED）感测的数据，发挥着重要作用。SCADA 子系统将作为联邦学习节点，使用各自的本地数据集训练本地模型，并向控制中心更新压缩梯度。此外，我们假设每个 SCADA 子系统均诚实且不具好奇心，因此彼此之间不会进行横向通信。

3)电网区域：电网区域被认为配备了对智能电网的分布式监测，该监测由传感器网络和IED（如继电器、同步相量测量装置等）组成。物理电网中的IED持续感知与电网相关的数据，包括电压相角、电流幅值、继电器状态等。假设这些IED通过高速通信网络连接到SCADA子系统，以传

频繁感知到的测量数据。

3.2. 提出的攻击检测模块

如图所示，电力系统中的网络攻击检测问题采用DAE模型进行建模2。DAE是功能强大的无监督表示学习神经网络，其目标是实现输入与输出之间的恒等映射[34]。自编码器通过将输入向量 x_i 传递给编码器 f_α 来学习重要特征，编码器将 x_j 映射为潜在空间表示（也常称为信息瓶颈），$y^{(j)}$ 该表示本质上是原始输入经过压缩后的低维知识表示，因此 $y^{(j)} = f_\alpha(x_j) = a(Zx_j + c)$，其中 a 是逐元素激活函数，Z 是编码器的权重矩阵，c 是编码器的偏置向量。随后，样本潜在向量 y_j 被传递给解码器 $g_{\alpha'}$，解码器通过 $x_j = g_{\alpha'}(y_i) = a(Z'y_j + c')$ 对原始输入 x_j 进行重构，同时尽可能降低重构误差。模型参数 Z、Z'、c 和 c' 将在联邦训练过程中不断调整，使得 α'、$\alpha' = \underset{\alpha, \alpha'}{\arg\min}\, \ell(x_j, y_j)$ 其中 $\ell(x_j, y_j)$ 是待最小化的重构误差函数。

所提出的DAE网络架构由两个对称的深度信念网络组成，前五个浅层构成网络的编码部分，后五个浅层构成模型的解码部分。这些层由受限玻尔兹曼机（RBM）组成，RBM是深度信念网络的基本构件。每个RBM层都支持其前一层与后一层之间的通信，即对于前一层节点而言，它充当隐藏层；

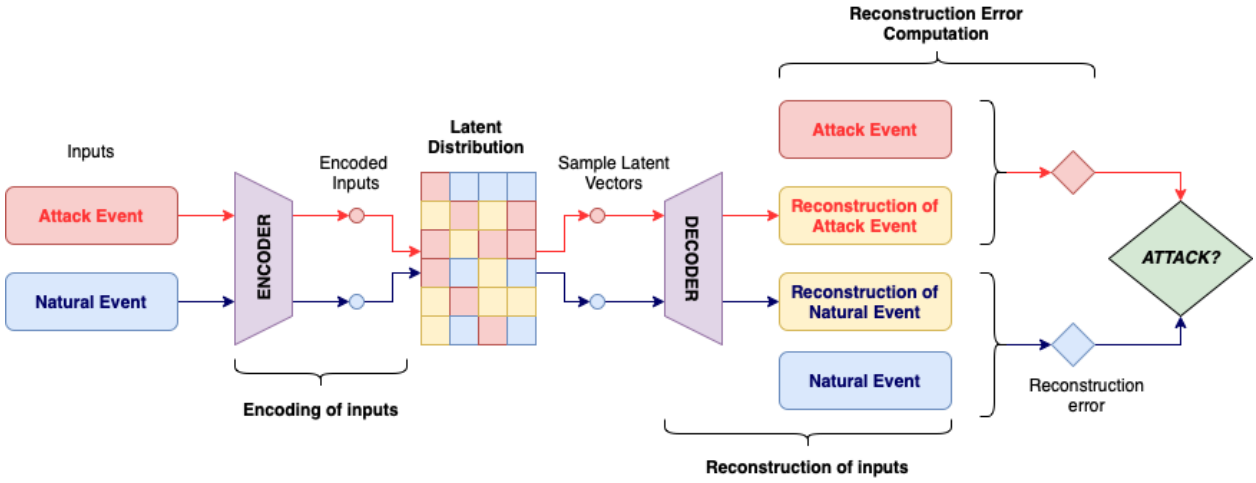


图 2：所提出的攻击检测模块架构示意图，并展示一个数据样本经过异常检测流程的示例。该检测模块由三个部分组成：1）编码器：编码器将输入压缩为维度更低的潜在空间表示。2）潜在分布：通过在网络中设置瓶颈，我们强制将原始输入表示为压缩的知识，使样本潜在向量能够输入解码器。3）解码器：解码器输出对原始输入的重构，随后将其与原始输入进行比较。

对于后一层节点而言，它充当可见层。然而，同一层节点之间不存在横向通信。RBM采用对比散度算法进行训练 [35]。此外，所有DAE层均采用修正线性单元（ReLU）激活函数，因为该函数能够提升RBM的性能，更易于训练，并且通常可以取得更好的效果。重构误差用于描述是否存在网络攻击。在本文场景中，所选用的重构误差指标为均方误差（MSE），其作为异常程度的度量。如果重构误差r超过某一阈值τ，则可能发生了网络攻击。在训练过程结束时，将训练集的重构误差按升序排列，并选取误差分布中出现拐点处的数值作为 t的值。为实现分类，我们在RBM堆栈的末端添加Softmax层。训练时的批次大小和学习率分别设置为100和 10^{-3} 。该模型构成了本文虚假数据注入攻击检测的基础。

3.3. 隐私保护梯度量化机制

在训练过程中，DAE 的目标是优化目标函数，$g(x) = \underset{\alpha,\ \alpha'}{\arg\min}\ \ell(x_j, y_j)$ 从而使重构误差最小化。通常，对于给定的任何恒等映射任务，$f : M \to N$，在FL 设置中，可以通过使用随机梯度下降（SGD）算法来搜索最优的 DAE 参数。该算法选择随机训练样本的损失来近似目标函数。然而，如图1所示，智能电网系统中的网络攻击检测大规模联邦训练需要在本地模型更新之间频繁进行双向通信控制中心和SCADA子系统。本地模型更新的大小增加了通信带-宽，从而限制了我们所提出的方法向更多客户端节点扩展的能力。此外，本地模型梯度的上下文可能会因易于访问而泄露隐私，因此不符合联邦学习 [11] 的隐私保护保障。

算法 1：梯度量化机制

输入：学习率 η，一位量化器 $sign(.)$, K 具有独立梯度的监控与数据采集子系统 g_k^t，当前权重参数 Z^t。

SCADA子系统：
计算归一化梯度 $\bar{g}_k^t = g_k^t - \mu_k^t$。
计算量化梯度 $\hat{g}_k^t = dpsign(\bar{g}_k^t)$。
发送 $\hat{g}_k^t$ 至控制中心。

控制中心：
获取 梯度 $\hat{g}_k^t$ 来自每个SCADA子系统 k.
通过多数投票计算新梯度 $\hat{g}_{MV}$
基于对所有 K （满足 $\hat{g}_{MV} =$）的 $\hat{g}_k^t$ 聚合

$$sign\left(\sum_{k=1}^{K} \hat{g}_k^t\right).$$

更新 新的模型参数 Z^{t+1} 使得
$Z^{t+1} = Z^t - \eta^t \hat{g}_{MV}$。
发送 Z^{t+1} 给每个 k。

因此，为了进一步提高我们所提出框架的通信效率，我们提出采用

梯度量化方法，通过将梯度近似为低精度值来降低通信带宽。具体而言，我们采用 DP-SIGNSGD [15] 算法：每个 SCADA 子系统在本地模型训练 k 后，对其本地梯度进行零均值归一化，g_k^t使得 $\bar{g}_k^t = g_k^t - \mu_k^t$ 所有 K 的一位量化输出呈均匀分布。每个 k 随后利用一位差分隐私量化器对其本地归一化梯度 $\bar{g}_k^t$ 进行压缩，该量化器仅取 $\bar{g}_k^t$ 的符号，使得 $\hat{g}_k^t = dpsign(\bar{g}_k^t)$。随后，每个 SCADA 子系统将其量化梯度 $\hat{g}_k^t$ 发送至控制中心，而非发送实际的本地梯度 g_k^t。另一方面，为获得一位压缩的全局梯度估计值，控制中心采用 [36]中详述的基于多数投票的聚合方法，对一位本地梯度之和取符号，使得 $\hat{g}_{MV} = sign\left(\sum_{k=1}^{K} \hat{g}_k^t\right)$。在估计出 $\hat{g}_{MV}$后，

控制中心按如下方式更新模型的新权重参数：$Z^{t+1} = Z^t - \eta^t \hat{g}_{MV}$其中，$\eta^t \in (0,1)$ 是 SGD 算法的学习率。该梯度量化技术结合多数投票进行模型更新聚合，如算法1 所示，可减少 SCADA 子系统与控制中心之间交换的梯度更新字节数，从而提高通信效率。

算法 2： 所提出的联邦学习框架

输入：学习率 η、一位量化器 $sign(.)$, K SCADA子系统、本地训练集 D_k for $k \in [1, K]$

初始化 共同一致的模型参数 Z_0.

对于 每个通信轮次 t *in* $T_{cl} \in (1, n)$ **执行**

控制中心:

获取 来自各个监控与数据采集子系统 k的梯度 。

通过 多数投票聚合，计算新的梯度 $\hat{g}_{MV}$

针对所有 $\hat{g}_k^t$ 进行 K，使得 $\hat{g}_{MV} = sign\left(\sum_{k=1}^{K} \hat{g}_k^t\right)$。

更新 新的模型参数 Z^{t+1} 使得 $Z^{t+1} = Z^t - \eta^t \hat{g}_{MV}$

发送 Z^{t+1} 给每个 k。

SCADA子系统:

对于每个 $k \in [K]$**执行**

计算本地梯度 $g_k^t = \nabla f_k(Z^t)$使用 D_k.

计算 归一化梯度 $\bar{g}_k^t = g_k^t - \mu_k^t$。

计算 量化梯度 $\hat{g}_k^t = dpsign(\bar{g}_k^t)$.

更新 $\hat{g}^t{}_k$ 到控制中心。

end

end

3.4. 联邦学习框架

如图所示 1，我们考虑一种联邦学习场景，其中一个控制中心连接到 K SCADA子系统，并通过通信系统（例如IEC 61850协议）进行通信。在我们提出的框架中，我们假设SCADA子系统与控制中心之间的数据通信协议是安全且可信的。联邦学习场景的主要目标是利用不同 SCADA子系统各自的本地数据集，协同训练全局去噪自编码器模型，而无需将敏感的电力相关数据上传至控制中心。每个 SCADA子系统的数据集均从相应电网系统区域内的多个 IED 中获取。设 $D_k = \{a_k^i, b_k^i\}_{i=1}^{N_k}$ 为特定 SCADA子系统的训练数据集 $k \in [K]$，其中 $N_k = |D_k|$是每个数据点的数量 D_k。需要注意的是，数据分布是非独立同分布的，这与现实世界中的网络攻击检测场景一致。因此，对于共享的全局 DAE 模型参数 f_k，目标函数$Z \in \mathbb{R}^M$可定义为 $f_k(Z) = \frac{1}{N_k}\sum_{i=1}^{N_k} \ell\left(\{a_k^i, b_k^i\}; Z\right)$，其中 $\ell(.): \mathbb{R}^M \times \mathbb{R} \to \mathbb{R}$ 是要优化的数据点 $\left(a_k^i, b_k^i\right)$的损失。此外，所有分布式 F 数据集上的全局目标函数 K 定义为 $F(Z) = \sum_{k=1}^{K} \frac{N_k}{N} f_k(Z)$，其为每个 k 的本地损失函数之和，其中 N 是用于训练的数据点总数。在特定的 t^{th} 通信轮次中，我们提出的

算法 2 中详述的联邦学习框架包含以下五个具体实现阶段：1）下行通信：控制中心将全局 DAE 模型参数 Z^t 分发给所有 K 个监控与数据采集子系统。为便于本研究，我们假设通信安全、延迟稳定且无错误，因此所有 K 个监控与数据采集子系统都能在大致相同的时间接收到正确的初始化模型参数 Z^t 。2）本地模型训练：每个监控与数据采集子系统k 针对 $k \in [K]$，使用其本地训练数据集 $D_k \in [K]$ 和 Z^t计算本地梯度 $g_k^t = \nabla f_k(Z^t)$。3）梯度量化：每个 $k \in [K]$ 对其本地梯度 g_k^t 进行零均值归一化，使得 $\bar{g}_k^t = g_k^t - \mu_k^t$。随后，按照第 3.3 节和算法 1 中的详细说明，使用 DP-SIGNSGD 对归一化梯度 $\bar{g}_k^t$ 进行量化，使得 $\hat{g}_k^t = dpsign(\bar{g}_k^t)$。4 ）上行通信：每个监控与数据采集子系统 k 针对 $k \in [K]$，通过无线通信协议将其本地计算得到的一位量化梯度发送至控制中心。5）模型聚合：控制中心从每个 $k \in [K]$接收到 $\hat{g}_k^t$ 后，执行基于多数投票的聚合，并对所有.....进行求和。

一比特 $\hat{g}_k^l$ 以获得全局梯度估计 $\hat{g}_{MV}$。控制中心随后将新的权重参数 Z^{l+1}更新至每个监控与数据采集子系统 $k \in [K]$。

4. 实验结果与讨论

各个SCADA子系统的网络攻击检测性能与我们所提出框架的整体性能直接相关，如第3节所述。因此，本节将从网络攻击检测性能和通信效率两方面评估我们所提出架构的性能。

4.1. 数据集与特征空间准备

我们使用密西西比州立大学和橡树岭国家实验室电力系统攻击（MSU-ORNLPS）数据集 [37]评估所提出的架构。该数据集包含128个特征和1个目标标记变量，数据日志由四个相量测量单元（PMU）采集
实时数字模拟器（RTDS）被用于在硬件在环测试台配置中模拟电力系统场景实施和数据生成。每个 PMU 以每秒 120 个采样点的速率采集实时同步相量测量数据，其中包括频率、电流相角、电流幅值、电压相角、电压幅值等。对于本研究，基于该数据集，我们假设攻击者能够渗透电网区域，并从每个电网区域的运行网络内部发起网络攻击。该数据集由 15 个数据文件组成，属于二分类数据，其中模拟了 A stem 扰动（标记为自然事件）和电力系统网络攻击（标记为 攻击事件）场景。网络攻击包括虚假数据注入攻击和跳闸命令注入攻击，具体如 [37]所述。

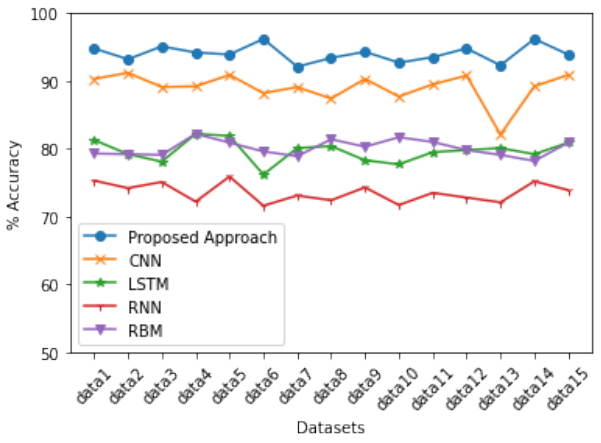


图3： 五种深度学习攻击检测模型在该数据集十五个数据文件上的平均性能比较。结果表明，在十五个数据文件的攻击检测方面，我们所提出的框架具有优越性。

.

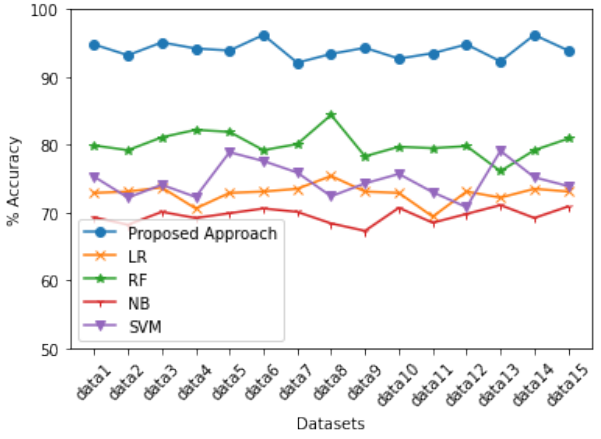


图4： 所提出的DAE模型与其他传统机器学习模型在该数据集十五个数据文件上的平均性能比较。同样，结果表明，在十五个数据文件的攻击检测方面，我们所提出的框架具有优越性。

为处理数据集中存在的缺失值（这可能是监控与数据采集通信故障所致[38]），我们采用传统的基于机器学习的 K 近邻（KNN）缺失值插补方法，以避免信息损失，而不是进行下采样。检测性能通常取决于特征空间基底的恰当选择。因此，我们应用主成分分析（PCA）算法，筛除一些难以重构的数据特征，使剩余特征能够通过压缩后的低维知识表示来表征，同时将重构误差降至最低。最终确定了由 100 个特征构成的基底。接下来，我们将输入数据归一化至 (0,1) 范围，以加快模型收敛。最后，我们抽取 30% 的数据集用于测试全局虚假数据注入检测模型的准确率，其余训练数据则分配至各个分布式电网区域，用于联邦学习本地模型训练。

4.2. 攻击检测性能

我们将所提出的FedDiSC框架在智能电网网络攻击检测中的攻击检测性能，与另外四种广泛使用的模型进行了评估，即卷积神经网络（CNN）、长短期记忆网络（LSTM）、循环神经网络（RNN）和受限玻尔兹曼机（RBM）。所有竞品模型均在相似的基于联邦学习的设置和配置下进行训练。我们在第4.1节所述的 MSU-ORNL数据集上评估这些模型。首先，我们在训练集上训练联邦学习模型，然后在MSU-ORNL PS测试集上对全局模型进行训练，并评估和比较其准确率。如图3所示，实验验证结果表明，所提出的 DAE模型在全部 15 个数据文件上的准确率均优于其他模型。例如，我们提出的 DAE模型在 data1 上的准确率为 94.8%，而 CNN 模型为 89.9%，LSTM 模型为 81.5%。同样，对于 data15，DAE模型的表现比 RBM 模型高 13.6%，比 RNN 模型高 19.9%。这表明，我们提出的 DAE模型对不同电网具有更强的鲁棒性——

学习模型 / 指标	所提方法	CNN	LSTM	RNN	RBM
准确率 (%)	94.8	90.3	81.3	75.3	79.3
精确率 (%)	92.7	91.8	78.5	73.5	78.2
召回率 (%)	92.8	90.9	78.1	73.0	77.7
F分数 (%)	93.6	91.1	79.2	73.2	77.9

表1
在整个数据集上，比较各种深度学习模型与我们提出的 DAE模型的平均攻击检测性能。我们将提出的 DAE模型与 CNN、LSTM、RNN 和 RBM 模型进行比较，这些模型均在联邦学习设置下训练。结果发现，在攻击检测性能方面，我们提出的方法优于所有竞品模型。

学习模型 / 指标	所提出的方法	LR	RF	NB	SVM
准确率 (%)	94.8	73.3	79.9	69.7	75.3
精确率 (%)	92.7	73.1	79.3	68.1	75.4
召回率 (%)	92.8	72.7	78.9	69.2	74.9
F分数 (%)	93.6	72.9	79.2	70.0	75.1

表 2
不同传统机器学习模型与我们提出的去噪自编码器在整个数据集上的平均攻击检测性能比较。我们将提出的联邦学习框架与集中式逻辑回归、随机森林、朴素贝叶斯和支持向量机等传统机器学习模型进行比较。结果表明，我们提出的框架在攻击检测方面更具优势。

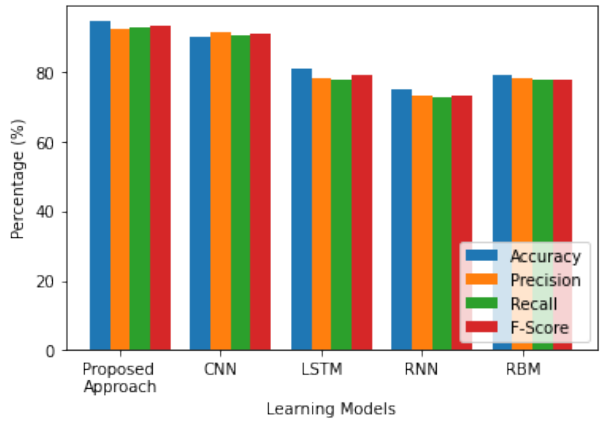


图 5： 基于准确率、精确率、召回率和 F分数指标，对整个数据集上另外四种深度学习攻击检测模型与我们提出的方法的平均性能进行比较。结果表明，在整个数据集上的攻击检测中，我们提出的框架具有优越性。

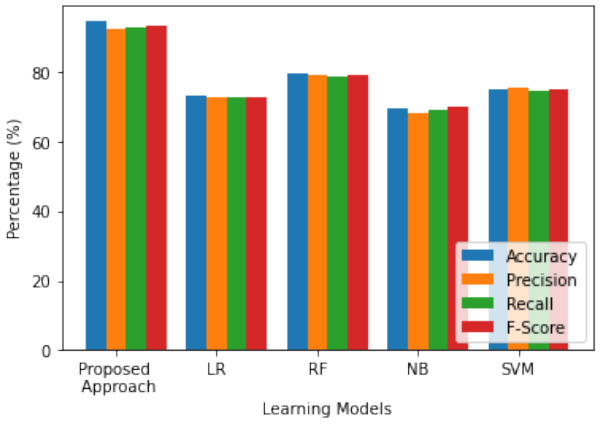


图 6： 基于准确率、精确率、召回率和 F分数指标，对整个数据集上四种传统机器学习攻击检测模型与我们提出的方法的平均性能进行比较。结果表明，在整个数据集上的攻击检测中，我们提出的框架具有优越性。

相关数据集，这是因为联邦学习能够从多个监控与数据采集子系统中学习特征，从而提升模型的鲁棒性。此外，我们使用四种分类指标，即准确率、精确率、召回率和 F分数，对全局去噪自编码器模型在测试集上的平均攻击性能进行了比较，结果如表1 和图 5所示。根据实验结果，我们观察到，所提出的模型在全部三项指标上均优于其他竞争模型。例如，DAE模型的精确率优于

分别比竞争模型 CNN、LSTM、RNN 和RBM 高 0.9%、14.2%、19.2%和 14.5%。同样，我们提出模型的召回率分别优于竞争模型 1.9%、14.7%、19.8% 和15.1%，而我们提出模型的F-Score分别优于上述模型 2.5%、14.4%、20.4% 和 15.7%。

此外，我们还将所提出方法的攻击检测性能与四种常用的传统机器学习模型进行了评估对比，这些模型曾用于

以往文献，即逻辑回归（LR）、随机森林（RF）、朴素贝叶斯（NB）和支持向量机（SVM）。在这些对比模型中，只有我们提出的模型采用联邦学习设置进行训练，其余模型均采用集中式训练。如图4所示，我们评估了所提出的全局去噪自编码器模型在全部 15 个数据文件上的准确率，并将其与四种不同的集中式传统机器学习模型进行了比较。例如，我们提出的模型在 data1 上的准确率为 94.8%，而 LR 模型为 72.9%，NB 模型为 69.4%。类似地，在 data 15 上，我们提出的模型准确率分别比 RF 和 SVM 高出 12.9% 和 10.0 %。实验结果如表2 和图6所示。我们发现，与其他竞争的传统机器学习模型相比，所提出的方法取得了最高的攻击检测性能。例如，与 LR、RF、NB 和 SVM 相比，我们提出的联邦学习方法精确率分别高出 19.6 %、13.4%、24.6% 和 17.3%。这可能是因为深度学习算法的性能优于传统机器学习模型：它们能够以渐进的方式从数据中学习高层特征，并随着数据集规模增大而不断提升。另一方面，我们的实验验证表明，在所有用于比较的模型中，朴素贝叶斯的整体性能最差，攻击检测准确率仅为 69.7%。

最后，我们将所提出的基于联邦学习的网络攻击检测框架的结果与该领域前沿文献中的结果进行比较。文献 [39] 中的作者开发了一种混合入侵检测系统，采用常见路径挖掘算法，并在类似的数据集上评估了所提出的方法。他们取得的总体平均准确率为 73.43%，而我们提出的框架准确率达到 94.8%。同样，与 [40],相比，我们的网络攻击检测准确率也有所提高。因此，可以得出结论：与其他深度学习模型相比，我们提出的用于网络攻击与电力系统扰动判别的联邦学习框架具有足够高的检测准确率，同时能够保障电网相关敏感信息的隐私，并具备较高的计算效率。

4.3. 所提方法的计算效率

随后，我们评估所实现的计算效率
如第3.3节所述，这得益于我们提出的虚假数据注入攻击检测框架所采用的梯度压缩方法。具体而言，我们比较了采用梯度量化（GQ）的所提出联邦学习框架与未采用 GQ 的传统联邦学习方法中模型的训练时间。为了评估所提出方法的计算效率，我们在两种设置中都采用了结构和超参数相近的模型进行比较。在图7中，我们比较了采用 GQ 方案的所提出联邦学习框架与未采用 GQ 方案的传统联邦学习方法的计算效率

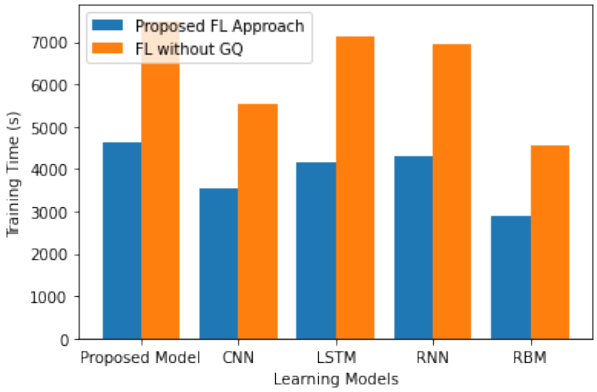


图7：采用梯度量化方案的联邦学习框架与未采用梯度量化方案的传统联邦学习方法之间的训练时间比较。结果表明，与其他方法相比，我们提出的框架可将训练时间缩短 40%。

在不同模型下未采用 GQ 方案的传统联邦学习方法。如图7所示，对于所有参与比较的模型，采用我们提出的 GQ 联邦学习框架后，模型训练时间均有所减少。具体而言，基于 GQ 的联邦学习方法的训练时间比未采用 GQ 的传统联邦学习框架约少 40%。这是因为，我们提出的联邦学习框架采用的隐私保护 GQ 方案将控制中心与监控与数据采集子系统之间共享的梯度大小压缩至 1 比特，同时不会影响模型的检测准确率，如第4.2节所述。

4.4. 研究结果总结

在本研究中，我们重点研究了智能电网中虚假数据注入攻击的检测，并假设该智能电网被划分为 N 个区域，其中 $N \in \mathbb{R}^{+}$。具体而言，我们提出了一种基于联邦学习的网络攻击检测方法，采用深度去噪自编码器模型进行电力系统扰动与网络攻击判别。我们使用 MSU-ORNL 电力系统数据集对所提出的方法进行了评估。如第 4.2节所示，与以类似方式训练的其他模型相比，我们提出的基于联邦学习的去噪自编码器方法在该数据集上实现了最佳的整体检测性能。此外，我们将该方法ver与集中式机器学习模型进行了比较，结果表明，所提出的框架在区分针对电力系统故障的网络攻击方面表现更优。同样，与最先进方法进行的简要比较显示出具有可比性的性能，从而验证了所提网络攻击检测方法在现实世界中的应用价值。此外，我们采用基于差分隐私 SIGNSGD 的梯度量化方案改进了所提出联邦学习架构的通信效率，该方案符合联邦学习保护隐私的特性。如图 7 所示，与传统的联邦学习检测模型相比，所提出框架的训练时间减少了 40%。因此，我们可以得出结论，所提出的框架能够提供充

分的检测准确率，同时提高虚假数据注入攻击检测的整体通信效率。

虽然模型性能足够出色是有效开展虚假数据注入攻击检测的重要指标，而且我们的工作已达到这一要求，但值得指出的是，我们提出的方法与以往工作显著不同，因为我们还考虑了电力相关敏感信息的安全与隐私。我们将模型训练限制在边缘侧的 SCADA 子系统上，并仅使用其本地数据集进行训练，从而无需向控制中心共享数据，也避免未经授权的攻击者访问原始数据。此外，采用隐私保护梯度量化方案并结合差分隐私，解决了联邦学习算法面临的严重深度梯度泄露问题。对手无法访问训练期间共享的模型原始梯度，因此无法提取电力相关敏感信息，甚至无法重建原始训练数据 [11]。因此，我们认为，所提出的隐私保护、通信高效框架适用于实际的智能电网攻击检测应用。

5. 结论

本文提出了一种新颖的通信高效且保护隐私的端侧联邦学习入侵检测方法，用于智能电网中电力系统扰动与网络攻击的辨别。首先，我们提出了一种基于联邦学习的框架，使电力传感设备能够协同训练入侵检测模型，而无需共享数据，从而解决数据隐私问题。其次，我们提出了一种基于表征学习的自动编码器，用于准确区分自然电力扰动与网络攻击事件。我们在一个常用的二分类电力系统数据集上，对所提模型的性能进行了全面评估，并将其与多种其他机器学习／深度学习算法进行了比较。实验结果表明，在该特定数据集上，自动编码器模型的准确率优于其他方法。最后，我们提出采用基于符号的梯度量化方案，以提升所提联邦学习框架的通信效率。实验结果表明，这种梯度量化机制可将训练时间缩短近 40%，证明量化梯度能够提升通信效率。

在未来的工作中，我们将利用集成学习方法，提高电力系统入侵检测模型的鲁棒性。此外，我们将重点研究提升联邦学习框架安全性的技术，因为我们此前的一项工作 [11]指出，联邦学习框架容易受到拜占庭攻击。

参考文献

[1] Muhammad Akbar Husnoo,Adnan Anwar, Nasser Hosseinzadeh, Shama Naz Islam, Abdun Naser Mahmood, and Robin Doss. 虚假主动配电系统中的数据注入威胁：一项综合调查，2021。[2] Muhammad Akbar Husnoo、Adnan Anwar、NasserHosseinzadeh、Shama Naz Islam、Abdun NaserMahmood和RobinDoss。Fe-drep：面向零售能源供应商的横向联邦负荷预测，2022。 [3] Po-Yu Chen、Shusen Yang、Julie A.McCann、Jie Lin和 Xinyu Yang。智能电网系统中的虚假数据注入攻击检测。*IEEE Communications Magazine*，53(2):206–213，2015。 [4] RaymondC. Borges Hink、Justin M. Beaver、Mark A. Buckner、Tommy Morris、UttamAdhikari 和 Shengyi Pan。用于电力系统扰动和网络攻击判别的机器学习。在 *2014 7th International Symposium on Resilient Control Systems (IS- RCS)*，第1–8 页，2014。 [5] Haftu Tasew Reda、AdnanAnwar 和 Abdun Mahmood。智能电网中虚假数据注入攻击的综合调查与分类：攻击模型、目标和影响。*Renewable and Sustainable Energy Reviews*，163:112423，2022 年 7 月。[6] AdnanAnwar 和 AbdunNaserMahmood。智能电网电力网络数据库中的异常检测：图匹配方法。*ElectricPower SystemsResearch*，133:51–62，2016年 4 月。[7] Mohammad Esmalifalak、Lanchao Liu、Nam Nguyen、RongZheng和Zhu Han。利用机器学习检测智能电网中的隐蔽虚假数据注入。*IEEE Systems Journal*，11 1652，2017 31644( ): – 。[8] Adnan Anwar、Abdun NaserMahmood 和 Zubair Shah。基于数据驱动的方法区分智能电网中的网络攻击与物理故障。在 *Proceedings of the 24th ACM International on Conference on Informationand Knowledge Management*，CIKM'15，第1811–1814 页，美国纽约州纽约市，2015。计算机协会。 [9] LeiWei、Donghuai Gao 和Cheng Luo。利用深度信念网络检测智能电网中的虚假数据注入攻击。在 *2018 Chinese Automation Congress (CAC)*，第 2621–2625页，2018。 [10] Thusitha Dayaratne、Mahsa Salehi、Carsten Rudolph和Ariel Liebman。面向智能电网安全分布式需求响应的虚假数据注入攻击检测。在 *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks(DSN)*，第 367–380 页，2022。[11] Muhammad Akbar Husnoo、Adnan Anwar、NasserHosseinzadeh、Shama Naz Islam、AbdunNaserMahmood和 Robin Doss。面向居民短期负荷预测的安全联邦学习框架，2022。 [12] Liang Zhao、JiamingLi、Qi Li和 Fangyu Li。用于检测太阳能电场中虚假数据注入攻击的联邦学习框架。*IEEE Transactions on Power Electronics*，37(3):2496–2501，2022。[13] Wen-Ting Lin、Guo Chen 和YuhanHuang。面向电网状态估计中虚假数据注入攻击检测的激励型边缘联邦学习：一种新颖的机制设计方法。*Applied Energy*，314，2022。 [14] Yang Li、Xinhao Wei、Yuanzheng Li、ZhaoyangDong和 Mohammad Shahidehpour。智能电网中的虚假数据注入攻击检测：一种安全的联邦深度学习方法。*IEEE Transactions onSmartGrid*，13(6):4862–4872，2022。[15] LingjuanLyu。Dp-signsgd：效率兼顾隐私性和鲁棒性，2021。[16] Seyed Hossein Majidi、Shahrzad Hadayeghparast 和 Hadis Karimipour。利用极端随机树算法和深度学习算法-自动编码器检测智能电网中的虚假数据注入攻击。*International Journal of Critical Infrastructure Protection*，37:100508，2022 年 7 月。 [17] Zubair Md. Fadlullah、Mostafa M. Fouda、Nei Kato、XueminShen和 Yousuke Nozaki。针对智能电网通信中恶意活动的预警系统。*IEEE Network*，25 55，20115 50( ): – 。[18] BM Ruhul Amin、AdnanAnwar和 M. J. Hossain。利用机器学习算法区分网络注入与故障。在 *2018 IEEE Region Ten Symposium (Tensymp)*，第 19–24 页，2018。 [19] Jinping Hao、Robert J. Piechocki、Dritan Kaleshi、Woon Hau Chin、

以及 Zhong Fan。智能电网中的稀疏恶意虚假数据注入攻击与防御机制。*IEEE* 工业信息学汇刊，11(5):1–12，2015。[20] Can Yang、Yong Wang、Yuhao Zhou、Jiongming Ruan 和 Wei Liu。使用机器学习方法检测电力系统中的虚假数据注入攻击。计算机与通信期刊，06(11):276–286，2018。[21] Ajit Kumar、Neetesh Saxena 和 Bong Jun Choi。用于检测电力系统中虚假数据注入攻击的机器学习算法。载于*2021* 年信息网络国际会议（*ICOIN*），第 385–390 页，2021 年。[22] Adnan Anwar、Abdun Mahmood、Biplob Ray、Md Apel Mahmud 和 Zahir Tari。利用机器学习确保电力系统拓扑网络数据库中的数据完整性。*Electronics*，9(4):693，2020 年 4 月。[23] Xiangyu Niu、Jiangnan Li、Jinyuan Sun 和 Kevin Tomsovic。利用深度学习动态检测智能电网中的虚假数据注入攻击。载于*2019* 年 *IEEE* 电力与能源学会创新智能电网技术会议（*ISGT*），第 1–6 页，2019 年。[24] Shuoyao Wang、Suzhi Bi 和 Ying-Jun Angela Zhang。智能电网中虚假数据注入攻击的定位检测：一种多标签分类方法。*IEEE* 物联网期刊，7(9):8218–8227，2020 年。[25] Yufeng Wang、Zhihao Zhang、Jianhua Ma 和 Qun Jin。Kfrnn：一种基于卡尔曼滤波器和循环神经网络的有效智能电网虚假数据注入攻击检测方法。*IEEE* 物联网期刊，9(9):6893–6904，2022 年。[26] Kübra Bitirgen 和 Ümmühan Başaran Filik。用于智能电网物理扰动与网络攻击检测判别的混合深度学习模型。关键基础设施保护国际期刊，40:100582，2023 年 3 月。[27] H. Brendan McMahan、Eider Moore、Daniel Ramage、Seth Hampson 和 Blaise Agüera y Arcas。利用去中心化数据进行深度网络的高通信效率学习，2016 年。[28] Afaf Taïk 和 Soumaya Cherkaoui。利用边缘计算和联邦学习进行电力负荷预测。载于*ICC 2020——2020* 年 *IEEE* 国际通信会议（*ICC*），第 1–6 页，2020 年。[29] Venkatesh Venkataramanan、Sridevi Kaza 和 Anuradha M. Annaswamy。利用保护隐私的联邦学习进行 DER 预测。*IEEE* 物联网期刊，第 1–1 页，2022 年。[30] Nastaran Gholizadeh 和 Petr Musilek。利用基于超参数的聚类进行电力负荷预测的联邦学习。物联网，17:100470，2022 年 3 月。[31] Yang Yang、Zijin Wang、Shangrui Zhao 和 Jinran Wu。一种用于短期负荷预测的集成联邦学习算法。电力系统研究，214:108830，2023 年 1 月。[32] Mi Wen、Rong Xie、Kejie Lu、Liangliang Wang 和 Kai Zhang。Fed-detect：一种用于智能电网窃电检测的新型隐私保护联邦学习框架。*IEEE* 物联网期刊，9(8):6069–6080，2022 年。[33] Muhammad Mansoor Ashraf、Muhammad Waqas、Ghulam Abbas、Thar Baker、Ziaul Haq Abbas 和 Hisham Alasmary。Feddp：一种利用联邦学习的隐私保护窃电检测方案。*Energies*，15(17):6241，2022 年 8 月。[34] Chong Zhou 和 Randy C. Paffenroth。利用鲁棒深度自编码器进行异常检测。KDD '17，第 665–674 页，美国纽约州纽约市，2017 年。计算机协会。[35] Sascha Lange 和 Martin Riedmiller。强化学习中的深度自编码神经网络。载于*2010* 年国际神经网络联合会议（*IJCNN*），第 1–8 页，2010 年。[36] Jeremy Bernstein、Yu-Xiang Wang、Kamyar Azizzadenesheli 和 Animashree Anandkumar。signSGD：用于非凸问题的压缩优化。载于 Jennifer Dy 和 Andreas Krause（编），第 *35* 届国际机器学习会议论文集，第 80 卷，第 560–569 页。PMLR，2018 年。[37] Uttam Adhikari、Shengyi Pan、Tommy Morris、Raymond Borges 和 Justin Beaver。[38] Adnan Anwar、Abdun Naser Mahmood 和 Mark Pickering。基于数据的智能电网不完整 *Measurements*，第 9650 卷，第 180–192 页。Springer International Publishing，Cham，2016 年。[39] Shengyi Pan、Thomas Morris 和 Uttam Adhikari。利用数据挖掘为电力系统开发混合入侵检测系统。*IEEE* 智能电网汇刊，6(6):3104–3113，2015 年。[40] Shengyi Pan、Thomas Morris 和 Uttam Adhikari。利用异构时间同步数据对电力系统中的扰动和网络攻击进行分类。*IEEE* 工业信息学汇刊，11(3):650–662，2015 年。