汽车充电桩分布式网络攻击聚类检测方法

吴 冉[1] 袁毕芳[1] 黄先宇[1] 章旭焱[1] 陈诗雨[1*]
WU Ran YUAN Bifang HUANG Xianyu ZHANG Xuyan CHEN Shiyu

摘 要 分布式汽车充电桩结构下，攻击行为分散在多个节点区域，单一维度的数据难以全面捕捉攻击特征，在依赖高维数据进行攻击检测时，高维数据往往具有稀疏性，降低了检测结果的准确性。因此，提出一种新的汽车充电桩分布式网络攻击聚类检测方法。提取平均字节数、平均数据包数、平均持续时间、源IP变化速率和端口变化速率等汽车充电桩分布式网络流量特征，引入K-means算法对汽车充电桩分布式网络流量特征进行聚类分析，成功检测出多种网络攻击类型。K-means算法通过综合考虑所有特征的距离，捕捉特征之间的潜在关系，从而弥补了多特征提取中因独立性导致的稀疏性问题。最后，利用异常得分设定阈值以区分正常流量与攻击行为。所提方法的攻击聚类结果在视觉上呈现出更为清晰的效果，对汽车充电桩分布式网络攻击的检测率达到95%，呈现出更为清晰的聚类效果。

关键词 汽车充电桩；分布式网络；网络攻击；聚类检测；K-means算法

doi：10.3969/j.issn.1672-9528.2025.07.020

0 引言

随着电动汽车的普及，汽车充电桩作为智能电网的重要组成部分，其安全性日益受到关注。充电桩网络通常采用分布式架构，以实现高效管理和资源调度，但这也使其成为网络攻击的主要目标。分布式网络攻击，如分布式拒绝服务（DDoS）、数据篡改和身份伪造等，不仅会破坏充电桩的正常运行，还可能威胁电网的稳定性和用户数据的安全。传统的攻击检测方法往往难以应对分布式网络中的复杂攻击模式，尤其是在多节点协同攻击场景下，检测效果有限。因此，开发一种针对汽车充电桩分布式网络攻击的聚类检测方法具有重要意义。

因此，针对攻击检测方法进行研究已经成为新能源汽车领域重点关注问题。文献[1]中针对网络攻击检测问题，设计一种基于FP-Growth算法的检测方法，运用FP-Growth方法进行历史数据中异常信号频繁模式的挖掘，并发现强有力的关联性，借助既有的频繁序列特征来辨识新出现的攻击类型及故障位置，从而达到检测网络攻击的目的。P-Growth算法主要关注频繁模式的挖掘，而对于稀疏特征的处理能力有限，无法有效利用稀疏特征中的信息。文献[2]中将CNN和LSTM结合在一起，进行网络攻击检测，运用标准化手段结合粒子群算法优化的K均值聚类（PSO-K-means）来处理含噪数据，并采用融合了卷积神经网络（CNN）与长短期记忆网络（LSTM）的多层级集成学习架构对噪声数据进行训练，以此增强分类器的精确度。CNN擅长处理局部相关性和空间特征，但对于高维稀疏数据，CNN无法有效提取有用的特征。文献[3]中构建了一个深度自编码器结构，开展网络攻击检测，深度自动编码架构能更有效捕捉数据的复杂特征，而序列到序列LSTM架构则擅长利用数据的时序属性，实现网络攻击的检测。LSTM擅长处理时间序列数据，但对于稀疏的时间序列数据，LSTM无法有效捕捉长期依赖关系，因为稀疏数据中的时间序列模式可能不够明显。

充电桩通常分布在不同的地理位置，且覆盖范围广泛，因此需要使用多维特征对其分布式网络攻击进行检测。然而，多维数据常展现出稀疏特征，这影响了检测结果的精确度。因此，开发一种适用于充电桩分布式网络的攻击聚类检测方法。K-means通过将数据点分配到K个簇中，将高维空间划分为多个低维子空间，每个簇内的数据点具有较高的相似性。这种划分降低了数据维度，使得稀疏性在局部范围内得到缓解。同时，K-means通过迭代优化簇中心和样本分配，增强了簇内数据的紧凑性，进一步减少了稀疏性的影响。

1 汽车充电桩分布式网络流量多特征提取

分布式网络环境下的汽车充电桩系统面临着多样化的攻击威胁，包括但不限于数据篡改、拒绝服务攻击及IP欺骗等。为了实现对汽车充电桩分布式网络攻击的有效检测，本次研究中首要任务在于精确提取网络流量中的多个关键特征[4]，这些特征能够为识别异常行为提供强有力的依据。

1. 国网湖北省电力有限公司襄阳供电公司 湖北襄阳 441002

（1）平均字节数，这是一个衡量数据流传输负载的重要指标，反映了在特定时间窗口内，每个数据流所传输的平均字节数量。具体计算公式为：

$$X_1=\frac{\sum_{j=1}^{J}Z_j}{J} \tag{1}$$

式中：X_1 表示汽车充电桩分布式网络流量数据流中平均字节数；Z_j 表示第 j 个数据流中数据包的字节数；J 表示分布式网络流量数据流总数。

（2）平均数据包数 [5]，该特征用于量化每个数据流中包含的数据包平均数量，具体计算公式为：

$$X_2=\frac{\sum_{j=1}^{J}S_j}{J} \tag{2}$$

式中：X_2 表示汽车充电桩分布式网络流量数据流中平均数据包数；S_j 表示第 j 个数据流中数据包数。

（3）平均持续时间，其描述了数据流的平均活跃时长，具体计算公式为：

$$X_3=\frac{\sum_{j=1}^{J}T_j}{J} \tag{3}$$

式中：X_3 表示汽车充电桩分布式网络流量数据流平均持续时间；T_j 表示第 j 个数据流的持续时间。

（4）源 IP 变化速率 [6]，反映了单位时间内源 IP 地址的变更频率，具体计算公式为：

$$X_4=\frac{A}{I} \tag{4}$$

式中：X_4 表示汽车充电桩分布式网络流量数据流源 IP 地址变化速率；A 表示数据流中源地址的数量；I 表示数据流的采样时间间隔。

（5）端口变化速率 [7]，用于衡量单位时间内端口使用的变化情况，具体计算公式为：

$$X_5=\frac{U}{I} \tag{5}$$

式中：X_5 表示汽车充电桩分布式网络流量数据流端口变化速率；U 表示数据流中端口数量。

通过上述公式，本文成功提取了平均字节数、平均数据包数、平均持续时间、源 IP 变化速率和端口变化速率等汽车充电桩分布式网络流量特征，为后续攻击检测提供基础数据。

2 聚类特征检测汽车充电桩分布式网络攻击

根据文中上述内容成功提取了汽车充电桩分布式网络流量特征之后，本文引入了 K-means 算法对提取的多项特征进行聚类分析，进而检测出攻击类型。在多特征提取过程中，特征之间存在较强的独立性。这种独立性导致在某些特征组合下，数据点非常稀疏，降低了攻击检测结果的可靠性。在聚类过程中，K-means 算法会综合考虑所有特征的距离，从而捕捉特征之间的潜在关系。通过聚类，独立特征之间的关系被隐含地建模，增强了特征之间的相关性。

从原始汽车充电桩分布式网络流量特征集 X_i（i = 1, 2, 3, 4, 5）中随机选取 k 个数据点作为初始聚类中心。在此过程中，衡量各汽车充电桩分布式网络流量特征点至 k 个聚类中心点的远近，依据最近原则归入相应聚类中心点所在的簇，共计形成 k 个簇，度量标准选用欧氏距离 [8]，具体计算公式为：

$$D(X,Y)=\sqrt{\sum_{i=1}^{n}(X_i-Y_i)^2} \tag{6}$$

式中：$D(X, Y)$ 表示汽车充电桩分布式网络流量特征点 X 与初始聚类中心 Y 之间的欧式聚类；X_i、Y_i 分别表示 X 和 Y 的第 i 个分量；n 表示分量数量。

根据上式即可求出原始汽车充电桩分布式网络流量特征集 X_i 中各数据点与初始聚类中心之间的距离，并将各个数据点划分至距离最近的簇中，完成初始聚类。然后，重新计算每个汽车充电桩分布式网络流量特征簇的聚类中心，即计算簇内所有数据点的均值 [9]：

$$Y_i^{'}=\frac{1}{|S_i|}\sum_{X_i\in S_i}X_i \tag{7}$$

式中：$Y_i^{'}$ 表示更新后的汽车充电桩分布式网络流量特征点聚类中心；S_i 表示第 i 个汽车充电桩分布式网络流量特征簇；$|S_i|$ 表示第 i 个簇内的汽车充电桩分布式网络流量特征点数量。

根据式（7）即可完成初始汽车充电桩分布式网络流量特征点聚类中心的更新。不断重复上述步骤，直到聚类中心不再发生变化 [10]，本文得到了多个聚类结果，每个聚类结果代表了一种网络流量模式，这些模式则对应着不同的汽车充电桩分布式网络攻击类型。

对于每个数据点，计算其与所属簇中心的距离作为异常得分：

$$m_i=d\left(X_i,Y_i^{'}\right) \tag{8}$$

根据异常得分的分布，设定一个阈值 T，用于判断是否为攻击：

$$\begin{cases}m_i>T,\text{攻击}\\ m_i\le T,\text{正常}\end{cases} \tag{9}$$

3 测试实验

3.1 实验准备

本章旨在通过对比实验，验证本文所设计汽车充电桩分布式网络攻击聚类检测方法的性能。汽车充电桩分布式网络如图 1 所示。实验中以基于 FP-Growth 算法的汽车充电桩分布式网络攻击检测方法和基于 CNN 和 LSTM 结合的汽车充电桩分布式网络攻击检测方法为对照组。基于汽车充电桩分

布式网络攻击检测的实验需求，按照表 1 所示数据配置实验环境参数。

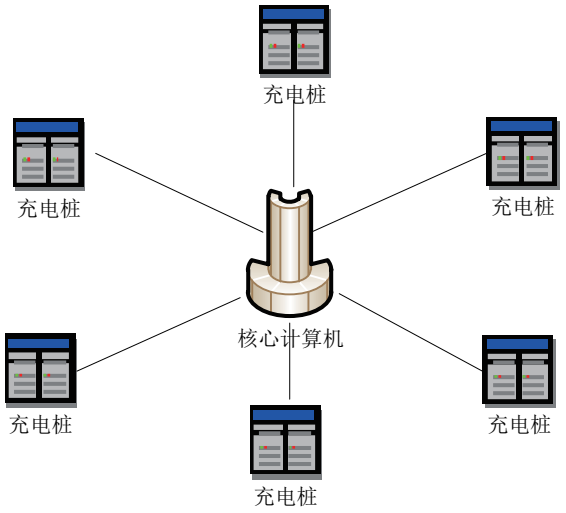


图 1 汽车充电桩分布式网络结构

表 1 实验环境参数配置表

实验环境	参数配置
硬件环境	Intel Xeon E5-2690 CPU, 128 GB RAM, NVIDIA Tesla K80 GPU
软件环境	Ubuntu 18.04 LTS, Python 3.7, TensorFlow 2.3, Scikit-learn 0.23
深度学习框架	TensorFlow
数据存储	MySQL 5.7

以 4 种常见类型的汽车充电桩分布式网络攻击为对象，分别收集一系列训练与测试样本，如表 2 所示。

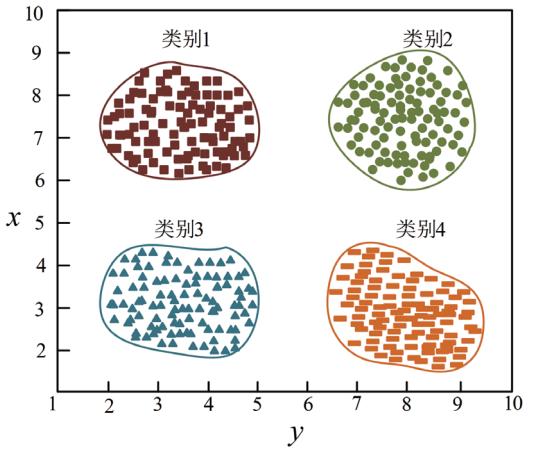
表 2 实验数据集分布表

类别标签	攻击类型	训练样本数	测试样本数
1	DOS 攻击	1 000	100
2	SQL 注入攻击	1 000	100
3	中间人攻击	1 000	100
4	数据篡改攻击	1 000	100

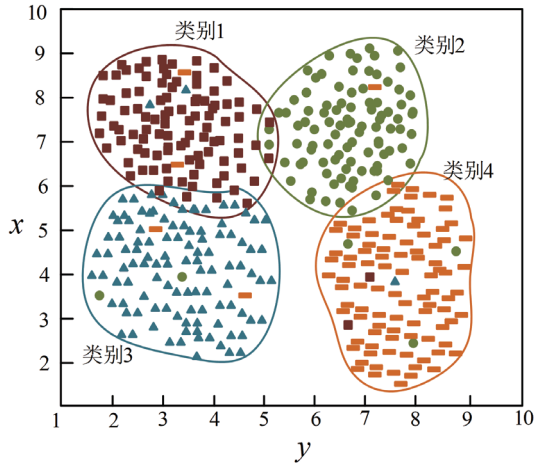
如表 2 所示，本次实验数据集分为训练和测试两个样本集。实验中，先应用训练样本集进行实验组方法和对照组中两种方法的训练；再分别采用训练完成的 3 种方法对测试集中各样本进行网络攻击检测，记录分析检测结果。

3.2 聚类效果分析

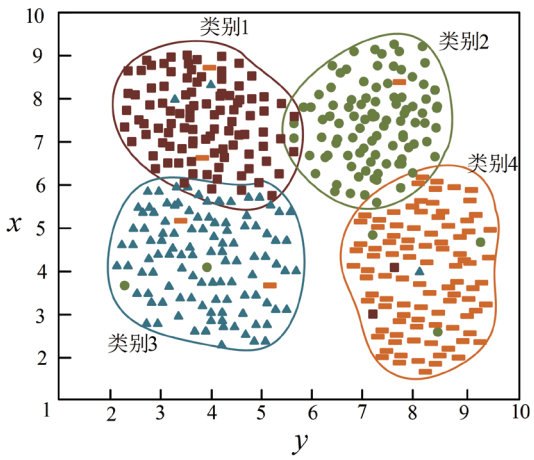
在本次实验中，为了更加直观地判断汽车充电桩分布式网络攻击检测效果，使用测试集中 4 种常见类型的网络攻击数据样本对 3 种方法进行了攻击检测后，记录各方法所得检测结果，并对这些结果进行可视化，如图 2 所示。



（a）所提出方法



（b）基于 FP-Growth 算法的检测方法



（c）基于 CNN 和 LSTM 结合的检测方法

图 2 分布式网络攻击聚类检测结果可视化图

如图 2 所示，与两种现有方法相比，本文方法所得汽车充电桩分布式网络攻击聚类结果在视觉上呈现出更为清晰的效果，4 种不同攻击类型下的汽车充电桩分布式网络流量特征样本均被较好地区分开来，且界限明显。但在两种对比方

法下，4 种不同攻击类型下的汽车充电桩分布式网络流量特征样本边界模糊，出现叠加，且大量样本被混淆。由此可以说明，本文研究的汽车充电桩分布式网络攻击聚类检测方法是有效且优越的。

3.3 攻击检测率

汽车充电桩作为智能电网的关键节点，面临多种网络攻击威胁，如数据篡改、身份伪造和分布式拒绝服务攻击等。攻击检测率测试能够量化方法在实际场景中对各类攻击的识别能力，确保其有效性和可靠性。通过测试，可以发现方法的薄弱环节，优化检测算法，提升对新型攻击的适应性。3 种方法的攻击检测率测试结果如图 3 所示。

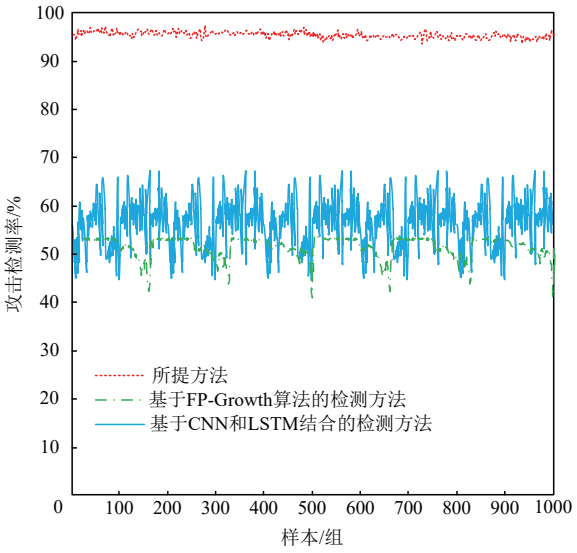


图 3 网络攻击聚类检测结果误报率

从图 3 数据可以看出，与两种现有方法相比，所提方法在汽车充电桩分布式网络攻击检测上表现出最佳的检测率。具体来说，在所提方法下，汽车充电桩分布式网络攻击检测结果的检测率达到 95%，较两种对比分别提升了 30%、40% 左右。因此，通过本次对比实验，验证了本文所设计汽车充电桩分布式网络攻击聚类检测方法具有良好检测性能。

4 结语

本文针对汽车充电桩分布式网络攻击检测问题，提出了一种基于流量特征的聚类检测方法。通过提取汽车充电桩分布式网络流量中的关键特征，并利用 K-means 聚类算法对特征进行聚类分析，实现了对不同类型的网络攻击的有效检测。实验结果表明，该方法相较于传统方法，在攻击检测性能上均有显著提升，能够有效应对充电桩分布式网络面临的多种信息安全威胁。本研究不仅为充电桩分布式网络的信息安全提供了一种有效的技术手段，也为电动汽车产业的健康发展提供了有力保障。未来，将进一步优化模型，提高检测效率，并探索更多应用场景，为充电桩分布式网络的信息安全提供更加全面、高效的解决方案。

参考文献：

[1] 李瑞，刘珊，闫磊．基于 FP-Growth 算法的新能源配电网 CPS 网络攻击检测方法 [J]. 电信科学 ,2024,40(11):103-113.

[2] 吕首琦，海涛，郑茂兴．基于 CNN 和 LSTM 结合的电网网络攻击检测（英文）[J]. 电子器件 ,2023,46(3):824-830.

[3] 黄燕，李金灿，杨霞琴，等．基于深度自编码器的智能电网窃电网络攻击异常检测 [J]. 电子技术应用 ,2024,50(2):76-82.

[4] 李元诚，罗昊，王欣煜，等．基于溯源图和注意力机制的 APT 攻击检测模型构建 [J]. 通信学报 ,2024,45(3):117-130.

[5] 周政雷，陈俊，潘俊涛，等．基于并行深度森林的配用电通信网络异常流量检测 [J]. 华东师范大学学报 (自然科学版), 2023(5): 122-134.

[6] 魏洪乾，时培成，张幽彤．汽车信息安全：面向总线网络的伪造攻击检测技术 [J]. 机械工程学报 ,2024,60(10):476-486.

[7] 陆鹏，付华，卢万杰．基于长短时记忆网络和生成对抗网络的 VRB 储能系统虚假数据注入攻击检测 [J]. 电网技术 , 2024, 48(1):383-393.

[8] 孙扬威，戚湧．基于聚类混合采样与 PSO-Stacking 的车载 CAN 入侵检测方法 [J]. 计算机工程 ,2023,49(1):138-145.

[9] 饶丹，时宏伟．基于深度聚类的航空交通流识别与异常检测研究 [J]. 计算机科学 ,2023,50(3):121-128.

[10] 宋世军，樊敏．基于谱聚类的多维数据集异常数据检测方法 [J]. 吉林大学学报 (工学版), 2023,53(10):2917-2922.

【作者简介】

吴冉（1987—），男，湖北荆门人，本科，高级工程师，研究方向：数据安全。

袁毕芳（1983—），女，湖北襄阳人，硕士，高级工程师，研究方向：信息技术。

黄先宇（1993—），男，湖北襄阳人，本科，工程师，研究方向：网络安全。

章旭焱（2000—），男，湖北襄阳人，本科，助理工程师，研究方向：网络安全。

陈诗雨（1997—），通信作者（email: chenshiyu2020@gmail.com），女，湖北襄阳人，硕士，助理工程师，研究方向：网络安全。

（收稿日期：2025-02-24 修回日期：2025-07-01）