Study on Intrusion Detection in IoTs Environment Based on GAN&CNN

LU Zhicheng[1], XU Haifeng[2], PAN Julong[1*]
(1.College of Information Engineering, China Jiliang University, Hangzhou Zhejiang 310018, China;
(2.School of Electronic Information, LiShui Vocational & Technical College, Lishui Zhejiang 323000, China)

Abstract: As technology advances, network intrusion techniques have become increasingly sophisticated, posing significant security challenges to edge devices within the Internet of Things (IoT) ecosystem. To overcome the limitations of traditional intrusion detection models in IoT environments——particularly their suboptimal detection performance and incompatibility with the resource-constrained nature of edge devices, a lightweight model that integrates generative adversarial networks (GANs) and convolutional neural networks (CNNs) is proposed for identifying intrusion behaviors in IoT settings. This innovative approach leverages GANs to address data imbalance issues, while a lightweight CNN based on the cross stage partial (CSP) structure is employed for efficient traffic feature extraction. Traffic data classification is achieved by using a Softmax function. Experimental results on the UNSW-NB15 and CICIDS2018 datasets demonstrate the model's exceptional performance, achieving accuracy rates of 99.64% and 99.65%, precision rates of 99.55% and 99.35%, recall rates of 99.61% and 99.64%, and F1 scores of 99.58% and 99.49%, respectively. Additionally, the model maintains a compact size of only 21 KB to 32 KB. These findings highlight the model's capability to deliver high-precision intrusion detection with minimal computational and storage requirements, making it an ideal solution for the demanding constraints of IoT environments.

Key words: intrusion detection; Internet of Things; generative adversarial network; convolutional neural network; cross stage partial structure
EEACC: 6150P; 7230 **doi**: 10.3969/j.issn.1004-1699.2025.10.017

基于 GAN&CNN 的物联网环境下入侵检测研究

卢志成[1], 徐海峰[2], 潘巨龙[1*]
(1.中国计量大学信息工程学院, 浙江 杭州 310018; 2.丽水职业技术学院电子信息学院, 浙江 丽水 323000)

摘 要: 随着科技进步, 网络入侵手段也越来越多样, 给物联网环境下边缘设备的安全带来了严峻挑战。针对目前物联网环境下传统入侵检测模型检测性能普遍较差以及不适配边缘设备资源受限、计算能力较低等特点, 提出一种基于生成对抗网络 GAN 和卷积神经网络 CNN 的轻量化模型用于检测物联网环境下的入侵行为。首先, 采用生成对抗网络技术解决数据不平衡问题; 其次, 使用基于跨阶段局部结构的轻量化卷积神经网络提取流量特征, 并选择 HSwish 作为激活函数, 以减少模型计算量和提高计算效率; 最后, 通过 Softmax 对流量数据进行分类。新算法在 UNSW-NB15 和 CICIDS2018 入侵检测数据集上进行实验, 模型检测入侵行为的准确率分别达到 99.64%和 96.65%, 精确率分别达到 99.55%和 99.35%, 召回率分别达到 99.61%和 99.64%, F1 分数分别达到 99.58%和 99.49%, 大小控制在 21 KB~32 KB 左右。结果表明, 所提出的模型在保证模型入侵检测精度的同时, 减少了模型的大小和计算量, 满足条件苛刻的物联网环境下高精度的入侵检测需求。

关键词: 入侵检测; 物联网; 生成对抗网络; 卷积神经网络; 跨阶段局部结构
中图分类号: TP183; TP212; TP393.08 **文献标识码**: A **文章编号**: 1004-1699(2025)10-1853-09

随着物联网(Internet of Things, IoTs)技术的迅速发展, 其应用范围不断扩展, 涵盖公共服务、车联网、智慧零售、智慧家居等多个领域, 为人们生活带来诸多便利。然而, 物联网的普及也带来了日益严峻的安全威胁。物联网设备之间的数据传输通常在开放的网络环境中进行, 因而易受异常流量攻击。这些异常流量可能表现为过量的数据请求、可疑的通信模式或未经授权的访问尝试等, 严重时可能导致设备功能失效或敏感信息泄露。因此, 构建有效的入侵检测系统(Intrusion Detection System, IDS)是维护物联网安全的关键。

目前, 国内外学术界在入侵检测方面已有大量的研究工作, 并取得了较好成果[1]。早期的入侵检测系统主要依赖规则和特征匹配, 这种方法在处理目前已知攻击时相当有效, 但对于新型的未知类型攻击的检测能力十分有限。随着机器学习技术发

收稿日期: 2024-10-10 修改日期: 2025-02-21

展,研究者们开始利用决策树[2]、支持向量机[3]、随机森林[4]等方法识别异常流量,这些方法能够通过学习正常流量模式来检测偏离正常行为的流量。但随着时代进步,网络攻击手段也越来越多样,这使得依赖人工设计并且不善于应对海量数据的基于机器学习入侵检测系统的检测效果越来越差。

近年来,深度学习的出现为入侵检测带来了新的机遇,学术界涌现出许多基于深度学习的入侵检测系统研究[5],如卷积神经网络[6](Convolutional Neural Network,CNN)、循环神经网络[7](Recurrent Neural Network,RNN)、自编码器[8](Autoencoder,AE)等。

然而,由于传统的入侵检测模型是部署在主机,无法适应物联网环境下的边缘节点资源受限和计算能力较低等特点,研究者们开始探索新的方法来优化入侵检测系统,以更好适应物联网安全需求。项睿涵等[9]提出一种一维卷积神经网络和差分隐私相结合的轻量化入侵检测模型,通过手动修改卷积通道数来达到减少模型参数量和轻量化的目的。孙红哲等[10]提出一种基于注意力机制和双向时间卷积神经网络(Bi-Directional Temporal Convolutional Network,BiTCN)的轻量化入侵检测模型,先通过使用 BiTCN 的轻量化结构减少模型参数量,再通过注意力机制动态分配不同输入部分重要性的特性提高模型特征提取效率来改善模型入侵检测性能。

为了进一步提高轻量化入侵检测模型的检测精度,研究者们使用过采样方法来平衡数据集,如 Zhang 等[11]提出一种结合了 SMOTE 方法(Synthetic Minority Over-Sampling Technique)和高斯混合模型(Gaussian Mixture Model,GMM)的过采样方法;Araujo-Filho 等[12]提出一种基于生成对抗网络(Generative Adversarial Network,GAN)的过采样方法。这些过采样方法通过平衡数据集的方式帮助模型更好地学习所有类别的特征,从而提升模型的入侵检测性能和鲁棒性。

综上所述,目前物联网环境中的入侵检测系统普遍存在模型对环境不适配和检测性能不高等问题。为解决这些问题,本文提出一种基于生成对抗网络和卷积神经网络的新型轻量化入侵检测模型,主要贡献有以下三点:①提出一种结合 WGAN-GP(Wasserstein GAN with Gradient Penalty)和 DCGAN(Deep Convolutional Generative Adversarial Network)优点的生成对抗网络模型 WDGP(Wasserstein Deep Convolutional GAN with Gradient Penalty)。WDGP 利用 WGAN-GP 的损失函数和梯度惩罚机制有效解决了传统 GAN 中训练不稳定和模式崩塌等问题,并通过 DCGAN 提供的深度卷积网络结构,使模型生成的数据质量显著提升,解决入侵检测训练中数据不平衡问题,提升了模型入侵检测的准确性和鲁棒性。②提出一种结合一维卷积神经网络(1DCNN)和跨阶段局部结构(Cross Stage Partial,CSP)优点的新型轻量化入侵检测模型 1DCSP。1DCSP 利用 1DCNN 高效的特征提取能力,使模型能快速提取入侵检测中流量的关键特征,然后通过 CSP 结构的引入,降低模型的参数量和计算复杂度,同时保持特征表达的完整性。这种轻量化设计使模型在资源受限的环境中能够以较低的计算成本实现高效的入侵检测。③模型在 UNSW-NB15[13]数据集上以 21 kB 的大小取得了 99.64% 的准确率、99.55% 的精确率、99.61% 的召回率和 99.58% 的 F1 分数;在 CIC-IDS2018[14]数据集上以 32 kB 的大小取得了 99.65% 的准确率、99.35% 的精确率、99.64% 的召回率和 99.49% 的 F1 分数。在 PC 上使用本文模型分别对两个数据集进行实验,模型在 128 个样本上的推理时间分别为 2.253 ms 和 20.390 ms。经实验证明,轻量化架构及实时性使得本文模型部署至物联网环境下边缘设备进行入侵检测成为可能。

1 相关背景知识介绍

1.1 数据集准备及预处理

本文实验在 UNSW-NB15 数据集和 CIC-IDS2018 数据集上进行,UNSW-NB15 数据集由包含 49 个特征和 10 种流量类型的数据构成,CIC-IDS2018 数据集由包含 78 个特征和 11 种流量类型的数据构成。

流量的原始特征数据需要经过清洗、贴标签和注释后才可作为模型输入,本文遵循以下步骤对数据集进行预处理。①数据特征缩减:在原始数据中存在一些无用特征,如 CIC-IDS2018 数据集中入侵者和用户都可以拥有的 Dst Port,Timestamp 等非必要特征数据,为了给数据特征降维,通过专家推理对数据特征进行删选;②异常值处理:在数据集中存在一些模型无法识别的数据,如 NAN、NULL 等,这些异常值会使模型无法正常训练,通过 DropNA 函数将这些异常数据删除;③数据标准化:不同特征存在不同的量纲和取值范围,为了避免因特征量纲不一致对模型分类造成的影响,首先通过 LabelEncoder 函数将数据流量中非数值型特征转换为数值型特征,以确保所有特征均为数字值,然后通过 StandardScaler 函数将数据标准化,消除量纲带来的影响。

1.2　基于 WDGP 和 1DCSP 的入侵检测模型

1.2.1　WDGP 介绍

生成对抗网络(GAN)是由 Goodfellow 等[15]在 2014 年提出的一种深度学习模型,主要用于生成与原始数据相似的高质量新数据。GAN 的核心思想是通过两个神经网络相互竞争和提升的方式来训练模型,这两个网络分别是生成器(Generator)和判别器(Discriminator)。生成器的目标是通过学习使生成的数据尽可能接近真实数据来“欺骗”判别器,判别器的目标是通过学习使其在判断真实数据为真的同时,尽可能地判断生成器生成的数据为假。具体过程如图 1 所示。

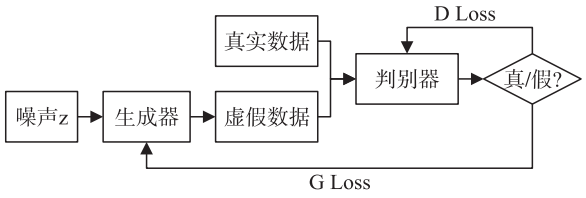


图 1　GAN 训练过程

图 1 中,D Loss 是判别器在真实数据上的损失和在生成数据上的损失之和,负责指导判别器进行优化;G Loss 是生成器生成的数据被判别器识别为真实数据的损失,负责指导生成器进行优化。

由图 1 可知,GAN 的训练过程可以看作是一个双玩家博弈,通过生成器和判别器相互对抗,交替优化不同的目标函数训练模型,最终达到使判别器无法区分真实数据和生成数据的目的。GAN 损失函数如式(1)所示:

$$\min_{G}\max_{D} V(D,G)=$$
$$E_{x\sim P_{\text{data}}(x)}[\log D(x)]+E_{z\sim P_z(z)}(\log\{1-D[G(z)]\}) \quad (1)$$

式中:D 表示判别器的损失,G 表示生成器的损失,$x\sim P_{\text{data}}(x)$表示从真实数据中采样,$z\sim P_z(z)$表示从生成器生成的虚假数据中采样,$D(x)$是判别器对真实数据 x 的输出,$D[G(z)]$是判别器对生成器生成的假数据 $G(z)$的输出。

WGAN[16](Wasserstein GAN)是在传统 GAN 的基础上提出的改进方法,旨在解决 GAN 训练中的不稳定性和模式崩塌等问题,它通过引入 Wasserstein 距离来改进损失函数,从而使训练更加稳定。相比于 GAN 使用的 JS 散度损失函数,WGAN 的 Wasserstein 距离损失函数提供了更直观的距离度量,尤其在数据不平衡和极端情况时,能帮助模型更有效地优化参数。在使用 Wasserstein 距离后,WGAN 的判别器输出不再是输出概率,而是一个评分函数,用来衡量样本的“真实程度”。在 WGAN 中的损失函数变为式(2):

$$L_C=E_{z\sim P_z(z)}\{C[G(z)]\}-E_{x\sim P_{\text{data}}(x)}[C(x)] \quad (2)$$

式中:$C(x)$是判别器对真实样本 x 的打分,$C[G(x)]$是判别器对生成样本的打分。由式(2)可知,WGAN 整体性能只需看判别器的分数,分数越高,生成器效果越好。

WGAN-GP[17]是对 WGAN 的重要改进,主要通过引入梯度惩罚来替代传统的权重剪裁,以进一步提高训练的稳定性和生成质量。相比于 WGAN 使用权重剪裁来保持判别器的 Lipschitz 连续性,即在每次更新后限制判别器的权重在某个范围内,WGAN-GP 引入了梯度惩罚机制,直接对判别器的梯度进行惩罚。这样做不仅能提高模型训练稳定性和生成数据质量,还能有效地保障 Lipschitz 的连续性。其损失函数变化如式(3)所示:

$$L_C=E_{z\sim P_z(z)}[C(G(z))]-E_{x\sim P_{\text{data}}(x)}[C(x)]+$$
$$\lambda E_{\hat{x}}[(\|\nabla D(\hat{x})\|_2-1)^2] \quad (3)$$

式中:$\hat{x}$ 是在真实数据和生成数据之间插值的样本,λ 是控制惩罚强度的参数。

本文在 WGAN-GP 基础上做进一步改进,提出了 WDGP 模型。WDGP 模型结合 WGAN-GP 和 DCGAN[18]的优点,既保留了 WGAN-GP 中判别器损失函数的变化,又融合了 DCGAN 中生成器和判别器架构上的变化,这样的好处是在提升生成器生成样本质量的同时,使生成对抗网络模型训练的过程更加稳定高效。WDGP 判别器的损失函数与 WGAN-GP 相同,如式(3)所示。WDGP 生成器和判别器的架构与 DCGAN 相同,相比于 WGAN-GP 中生成器和判别器全连接的架构,WDGP 在生成器中加入卷积操作帮助生成器能够生成具有空间结构和细节的流量特征,这样的特征数据更符合真实流量特征;在判别器中加入卷积操作提升判别器的特征提取能力,帮助判别器更好地区分真实样本和生成样本,同时由于判别器性能的提升,也能帮助生成器避免训练过程中的过拟合。WDGP 生成器和判别器的具体架构如表 1 和表 2 所示。

表 1　WDGP 生成器架构

层	卷积核	输出	激活函数	参数量
input	—	128×8×200	—	0
Conv	7×7	128×4×194	LeakyReLU	228
Conv	5×5	128×2×190	LeakyReLU	42
Conv	3×3	128×1×188	LeakyReLU	7
FC	—	128×1×66	Tanh	12 870
output	—	128×1×66	—	—

表 2 WDGP 判别器架构

层	卷积核	输出	激活函数	参数量
input	—	128×1×66	—	0
Conv	3×3	128×2×194	LeakyReLU	8
Conv	5×5	128×4×190	LeakyReLU	44
Conv	7×7	128×8×188	LeakyReLU	232
FC	—	128×200	LeakyReLU	86 600
FC	—	128×1	—	201
output	—	128×1	—	—

在表 1 生成器的架构中,使用了三层卷积和一层全连接实现对流量数据的特征生成。其输入为 128 个随机噪声向量,随机噪声向量的维度为 8×200。这样取高纬度的噪声向量好处是允许生成器能够探索和生成更加多样化的特征,如果噪声向量的维度过低,则生成器在映射到真实数据的过程中可能受到限制,无法捕捉到真实数据的复杂性。通过三层由大到小卷积核的卷积操作放大噪声向量,再使用一层全连接层将放大后低维噪声向量映射到高维的向量中,从而模拟真实流量数据。在卷积操作中使用 LeakyReLU 激活函数而不是 ReLU 函数可以有效避免 ReLU 在某些情况下导致的“死神经元”问题,并且允许网络在负半轴有一小部分梯度传递,从而防止梯度消失的问题。在最后的全连接层中使用 Tanh 激活函数,确保生成器生成流量样本数据与现实中归一化后的流量数据相同。在表 2 判别器的架构中,使用了三层卷积和两层全连接层对流量数据进行判别,其输入为 128 个流量特征向量。对应生成器的卷积操作,判别器通过三层由小到大卷积核的卷积操作充分提取输入数据的特征,再使用两层全连接层输出结果。由于 WDGP 采用的是与 WGAN-GP 判别器相同的判别方式,故与传统判别器在最后全连接层使用 Sigmoid 激活函数不同,它不需要输出[0,1]间的概率,而是输出一个用来评价数据真实度的实数,故最后使用线性激活来实现,即不改变输入数据的值,直接返回输入数据本身。

2.2.2 1DCSP 介绍

CSPNet(Cross Stage Partial Network)是一种用于提升卷积神经网络效率和性能的深度学习架构。它最早由 Wang 等[19]在 2020 年提出,旨在优化网络的计算性能、降低模型复杂度,同时保持高精度。CSPNet 的主要思想是通过跨阶段局部融合特征来提升模型的学习能力,并减少冗余计算。具体结构如图 2 所示。

图 2 中,Conv 主要通过卷积操作提取输入数据中的重要特征;Copy 负责复制输入数据,保留原始信息,避免在卷积操作中丢失关键信息;Concat 负责将经过卷积处理的特征图与复制的原始特征图在通道维度上进行拼接,实现特征融合。

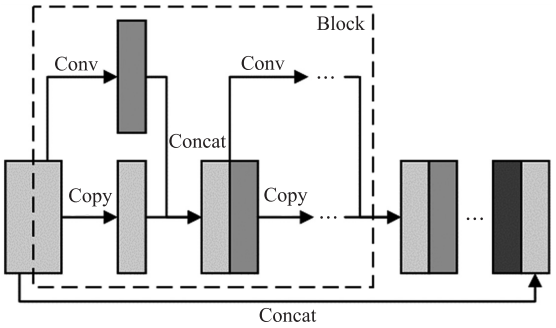


图 2 CSP 结构

由图 2 可知,CSP 结构通过将输入特征图分割为两部分,一部分通过 Copy 直接输出到结果,另一部分在设置的 Block 中进行卷积操作来提取特征,特征提取完成后再将结果与另一部分数据在通道上进行合并,从而实现特征的融合。

HSwish 是由 Google 团队在其 MobileNetV3[20]架构中提出的新型激活函数,它被设计用于轻量级神经网络,以提高模型的性能和计算效率,尤其适合物联网和边缘计算场景。HSwish 计算公式为:

$$\mathrm{HSwish}(x)=x\frac{\mathrm{ReLU}(x+3)}{6} \tag{4}$$

相比于 ReLU 公式中输入为负值则直接转化为 0 的非连续性特性,HSwish 结合了线性与非线性的特性,它的设计使其在输入值为负时表现得像 ReLU,而在正值区间时则提供更丰富的特征信息,是平滑的激活函数。这样的设计既可以减轻梯度消失问题,又优化了特征表示能力,允许更多的信息通过,提高了模型的训练和推理速度。

本文在 1DCNN 的基础上结合 CSP 和 HSwish 的特点,提出 1DCSP 模块,其结构如图 3 所示。

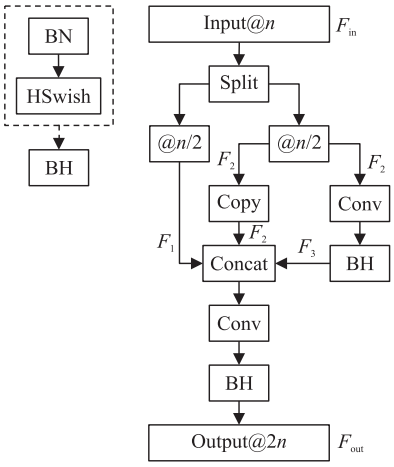


图 3 1DCSP 模块结构

图 3 中,Split 用于分割输入数据通道;BN(Batch Normalization)是批量归一化,用于加速训练和提高模型稳定性;激活函数 HSwish 用于帮助模型学习复杂的特征和模式,同时优化计算效率;@ n 代表数据当前通道数为 n;BH 是 BN 和 HSwish 的组合。

在选择分割通道数量时,浅层特征通道过少会导致信息丢失,深层通道过多则可能导致计算量过大。根据入侵检测任务的复杂性和计算资源,1DCSP 模块将输入特征 F_{in} 通道按照[50%,50%]的比例分割成两部分[F_1,F_2],其中 F_1 是保存了输入特征 F_{in} 原始信息的浅层特征,包含了原始信息中简单模式下的基础特征信息,确保在提取深层特征过程中不丢失重要的信息;F_2 与 F_1 内容相似功能不同,它是提取深层特征信息的基准,为之后深层特征提取提供参考。在成功划分通道后,对 F_2 进行局部特征提取操作,首先通过 Copy 操作保留未修改的特征信息,然后通过卷积核为 3 的卷积 Conv 操作和批量归一化 BH 操作,从 F_2 中提取更深层次的特征 F_3,F_3 是从原始信息中通过卷积和激活处理后得到的深层特征,能帮助模型更好地理解数据中的高级模式,尤其是复杂任务中需要捕捉的深层次关联。之后,F_1、F_2 和 F_3 通过 Concat 操作根据通道拼接浅层和深层特征,拼接后的特征保留了浅层信息和深层信息,使模型能够同时关注到低细节和高层抽象内容。最后,为实现跨阶段局部的特征融合,通过将拼接后的特征经过卷积核为 1 的 Conv 操作和 BH 操作处理,帮助模型有效地整合来自不同层次的信息,得到实现跨阶段局部融合的特征 F_{out},具体操作如式(5)和式(6)所示:

$$F_3 = \mathrm{BH}[\mathrm{Conv}(F_2)] \tag{5}$$

$$F_{out} = \mathrm{BH}\{\mathrm{Conv}[\mathrm{Concat}(F_1, F_2, F_3)]\} \tag{6}$$

通过跨阶段局部融合特征方式和 HSwish 激活函数,1DCSP 模块能有效提升模型的特征学习能力,同时减少模型训练时所需计算的参数量,从而达到轻量化模型和改善模型检测性能的目的。

1DCSP 模型具体结构如图 4 所示,由 1DCNN、1DCSP、Maxpool1D 和 FC 组成。其中 1DCNN 和 1DCSP 负责提取特征;Maxpool1D 是最大池化层,负责降低特征尺寸和提取关键特征;FC 是全连接层,负责整合特征和预测结果。

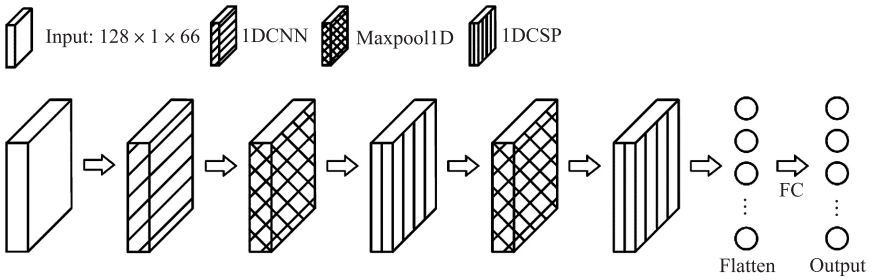


图 4　1DCSP 模型结构

1.3　模型训练具体过程

本文提出基于 GAN 和 CNN 的入侵检测模型 WDGP-1DCSP,训练具体过程如图 5 所示,包括数据预处理、模型训练和模型测试三个阶段。①数据预处理,包括无用特征缩减、异常值处理、数据标准化;②模型训练,包括数据集划分、WDGP 模型训练及生成样本、1DCSP 模型训练;③模型测试,训练好的模型在测试集上进行测试并分析测试结果。

首先,原始特征数据经过预处理成为适合模型处理的输入数据。然后,通过统计各类型流量数据,将需要进行过采样的少数类样本放入 WDGP 中进行训练。WDGP 主要由两个模型构成,分别是生成器和判别器,生成器以卷积-全连接的方式构成,输入是经过重构的噪声数据,输出是经过标准归一化

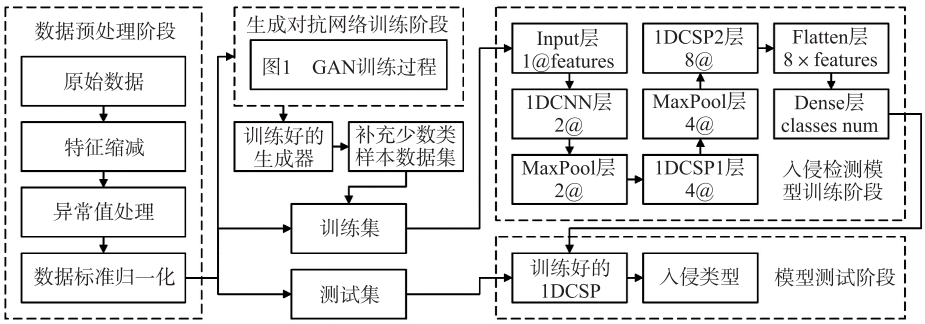


图 5　WDGP-1DCSP 架构

的特征数据;判别器以卷积-全连接的方式构成,输入是经过标准归一化的特征数据,输出是一个实数值,该值是对样本真实度的评分,判断数据是否真实。在 WDGP 训练过程中,它的生成器和判别器通过对抗的方式提升各自的性能,等到生成器可以欺骗判别器的时候,将训练好的生成器模型保存作为生成数据的工具。将数据集按照 8∶2 比例划分为训练集和测试集之后,使用保存的生成器模型生成少数类样本数据平衡训练集,再把训练集放入 1DCSP 模型中训练。1DCSP 模型以 1DCNN-池化-1DCSP-全连接的方式构成,输入为经过标准归一化的特征数据,输出为模型判断的流量类型。由于输入数据是一个通道,而 1DCSP 模块至少需要两个通道才能划分,故先通过 1DCNN 对数据进行初步处理,在提取特征的同时生成两个通道以供后续 1DCSP 模块的使用。在 1DCSP 模块中,输入数据经过图 3 所示跨阶段局部融合特征方式获得更丰富的特征表示,这种方式不仅有效减少了模型的计算量,还增强了模型提取特征的能力,使模型在处理复杂的流量数据时,能更好地捕捉不同层次的特征信息,从而帮助模型轻量化和改善入侵检测性能。最后,在测试集上测试训练好的 1DCSP 模型,并对实验结果进行分析。

2 实验及分析

2.1 实验设置及评价指标

本文实验在 64 位的 Intel® Core™ i7-12700H CPU@ 2.30GHz 计算机上进行,该机配有 16 GB RAM 和基于 Python 的 Nvidia GeForce GTX 3060 GPU(16 GB),使用 Python 的 Pytorch 库进行本文模型的编写,模型相关参数如表 3 所示。

表 3 模型相关参数

模型参数	参数内容
epoch	50
批处理大小	128
损失函数	CrossEntropyLoss
优化器	Adam
优化器学习率	0.005
Dropout	0.5
数据集划分比例 n_1/n_2	0.2/0.8

对于一个分类问题,其分类结果如表 4 所示。

表 4 分类结果

分类	预测正类	预测负类
实际正类	TP	FN
实际负类	FP	TN

评价指标分别是准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数(F1-score),如式(7)~(10)所示:

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{7}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{8}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{9}$$

$$\text{F1-score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{10}$$

2.2 实验结果及分析

入侵检测实验分为二分类(有、无入侵)和多分类入侵。

2.2.1 二分类入侵模型对比实验

为了证明 1DCSP 在轻量化二分类入侵检测中具有较好的检测性能,本文选择在 UNSW-NB15 和 CICIDS2018 上进行多个模型的对比实验,对比的轻量化模型包括 MLP[5]、1DCNN[9]、Attention-BiTCN[10]、MobileNet[20]、LSTM-ResNet[21]、LNN[22] 和本文提出的 1DCSP。从图 5 可知,1DCSP 最后输出为八个通道,为了更准确评估模型结构本身的优劣,其余对比模型在保证整体框架不变的情况下,输出通道也改为八个通道,其中特例为 MLP,因为是由全连接构成,故调整其隐藏层数量来保证参数量与 1DCSP 尽量相同。在对比实验中,所有模型最后的全连接层前均使用 Dropout 来增强模型的鲁棒性和泛化性。对比实验结果如表 5 和表 6 所示。

表 5 基于 UNSW-NB15 的入侵检测结果

入侵检测模型	评价指标/%			
	准确率	精确率	召回率	F_1 分数
MLP[5]	97.24	97.12	97.48	97.30
1DCNN[9]	98.27	98.11	98.38	98.24
Attention-BiTCN[10]	97.57	97.44	97.74	97.58
MobileNet[20]	99.24	99.25	99.20	99.23
LSTM-ResNet[21]	96.31	96.10	96.58	96.34
LNN[22]	95.89	95.69	95.79	95.74
1DCSP	**99.64**	**99.55**	**99.61**	**99.58**

表 6 基于 CICIDS2018 的入侵检测结果

入侵检测模型	评价指标/%			
	准确率	精确率	召回率	F_1 分数
MLP[5]	99.61	99.21	99.53	99.36
1DCNN[9]	99.18	99.23	98.78	98.99
Attention-BiTCN[10]	99.41	99.33	99.12	99.22
MobileNet[20]	99.49	99.18	99.43	99.30
LSTM-ResNet[21]	99.52	99.33	99.39	99.36
LNN[22]	99.15	99.13	98.10	98.61
1DCSP	**99.65**	**99.35**	**99.64**	**99.49**

分析表 5 和表 6 可得,本文所提 1DCSP 模型在不同数据集的四项指标上均取得较好的效果,在 UNSW-NB15 和 CICIDS2018 分别达到 99.64% 和 99.65%的准确率,证明 1DCSP 能够正确识别大部分流量,同时,在 UNSW-NB15 和 CICIDS2018 分别取得 99.58 和 99.49 的 F1 分数,证明 1DCSP 漏报和误报流量数量较少,具有较好的鲁棒性和泛化性,总体上 1DCSP 相比其他模型具有更好的入侵检测性能。

2.2.2　轻量化对比实验

为了进一步证明 1DCSP 具有较好的轻量性能并可以适应物联网环境下的边缘设备条件,本文选择在 UNSW-NB15 和 CICIDS2018 上进行多个模型大小和推理时间的对比实验,其中推理时间是模型推理一个 epoch 数据的时间。轻量化对比实验结果如表 7 所示。

表 7　模型大小(kB)及推理时间　单位:ms

入侵检测模型	数据集			
	UNSW-NB15		CICIDS2018	
	大小	时间	大小	时间
MLP[5]	58	**0.785**	69	**18.588**
1DCNN[9]	34	1.131	46	18.678
Attention-BiTCN[10]	129	1.294	282	18.742
MobileNet[20]	27	1.342	39	22.382
LSTM-ResNet[21]	137	2.196	290	19.913
LNN[22]	42	1.628	54	19.787
1DCSP	**21**	2.253	**32**	20.390

分析表 7 可得,1DCSP 模型在不同数据集上均表现了较好的轻量化性能,适合部署于物联网环境下的边缘设备上。与其他模型相比,1DCSP 在模型大小上具有较显著优势,尽管 1DCSP 的推理时间不是最佳,但其仍保持在可接受的范围内。

2.2.3　多分类入侵模型对比实验

为了证明 WDGP 可以提高物联网环境下轻量化入侵检测多分类模型的入侵检测性能,本文选择在 CICIDS2018 数据集上进行多个过采样方法的对比实验,过采样方法包括 SMOTE[11]、GAN[12]和本文提出的 WDGP。在开始对比实验前,先统计 CICIDS2018 数据集中的少数类样本以便于数据集的过采样。CICIDS2018 数据集有 11 种不同类型数据,有 5 种少数类数据,分别是 Bot、Web、XSS、LOIC-UDP 和 Slowloris,其分布如表 8 所示。

其中,Bot 攻击通过发送大量的流量占用攻击对象的服务器资源,达到用户无法正常获得服务的目的,它的攻击流量特点为短时间内来自同一 IP 或几个 IP 发送的大量字段请求;Web 和 XSS 攻击通过发送大量具有注入式特点的流量,达到攻击者绕过系统获取用户信息的目的,它们的攻击流量特点为大量且请求参数包含恶意代码或特殊字符;LOIC-UDP 攻击与 Bot 攻击相似,也是发送大量的重复无用流量,不过 Bot 一般针对应用层,而 LOIC-UDP 直接针对网络层,它的攻击流量特点为随机 IP 发送的大量简单无用的 UDP 数据包;Slowloris 攻击通过发送不完整的请求即低速流量来耗尽攻击对象服务器资源,它的攻击流量特点为数据包发送缓慢且连接时间长。

表 8　CICIDS2018 少数类样本数据分布

类别	训练集	测试集
Bot	16 690	4 102
Web	502	109
XSS	184	46
LOIC-UDP	1 391	339
Slowloris	8 791	2 199

分析表 8 可得,CICIDS2018 数据种类分布是不均匀的,相比于训练集拥有 86 523 条多数类样本 Benign 类型数据,表 8 少数类样本中最少的 XSS 类型只有 Benign 类型的 0.21%,这会导致模型倾向预测占多数的类,而忽视少数类,从而影响模型的整体的预测效果。分别对正常数据集、SMOTE[11]过采样数据集、GAN[12]过采样数据集和 WDGP 过采样数据集进行 1DCSP 的多分类入侵检测实验,实验结果如表 9~表 12 所示。

表 9　1DCSP 多分类入侵检测结果

过采样方法	评价指标/%			
	准确率	精确率	召回率	F_1 分数
正常	93.52	94.43	85.00	89.47
SMOTE[11]	99.08	87.15	**95.69**	91.22
GAN[12]	99.44	99.06	91.74	95.20
WDGP	**99.55**	**99.54**	91.68	**95.75**

表 10　1DCSP 多分类少数类样本入侵检测精确率

过采样方法	类别				
	Bot	Web	XSS	LOIC-UDP	Slowloris
正常	**100.00**	89.87	**100.00**	**100.00**	**99.13**
SMOTE[11]	95.11	29.51	2.12	85.13	95.81
GAN[12]	97.78	88.42	70.73	**100.00**	97.33
WDGP	98.65	**98.68**	63.79	**100.00**	98.96

表 11　1DCSP 多分类少数类样本入侵检测召回率

过采样方法	类别				
	Bot	Web	XSS	LOIC-UDP	Slowloris
正常	49.14	65.13	50.00	99.12	98.95
SMOTE[11]	48.82	**77.98**	**91.30**	99.71	**100.00**
GAN[12]	94.90	77.06	63.04	**100.00**	99.82
WDGP	**99.98**	68.81	84.09	**100.00**	99.59

表 12 1DCSP 多分类少数类样本入侵检测 F1 分数

过采样方法	类别				
	Bot	Web	XSS	LOIC-UDP	Slowloris
正常	65.89	75.52	66.67	99.56	99.03
SMOTE[11]	64.52	42.81	4.14	91.46	97.86
GAN[12]	96.31	**82.35**	66.66	**100.00**	98.55
WDGP	**99.31**	81.08	**72.55**	**100.00**	**99.28**

表 9 中所示为 1DCSP 在三种过采样方法下多分类入侵检测的结果。除在 SMOTE 过采样数据集的入侵检测精确率上比未经过过采样的正常数据集入侵检测精确率低,正常数据集其他入侵检测评价指标都要低于另外三种经过过采样方法后数据集的入侵检测评价指标,说明三种过采样方法都可以有效提高模型在多分类情况下的入侵检测性能。三种过采样方法下的入侵检测准确率相差不大,这是数据不平衡带来的影响,由于在测试集中多数类数据远多于少数类数据,故不能作为参考。SMOTE 由于其本身是在两个最近样本间插值方式形成生成样本,导致生成的少数类样本在正常样本之间,使模型学习到错误的特征,把不属于少数类样本的特征误认为少数类样本的特征,容易误报的同时不容易漏报,故其精确率最低而召回率最高。GAN 由于并未在生成器和判别器中使用卷积操作,故其生成的样本特征相较于 WDGP 略差一点,在准确率、精确率和 F1 分数上也略低于 WDGP。WDGP 在准确率、精确率和 F1 分数方面,均为最高,证明其过采样后数据集的多分类入侵检测性能综合表现在三种过采样方法中最好。表 10~表 12 所示为 1DCSP 在三种过采样方法下多分类中少数类样本的入侵检测结果。LOIC-UDP 和 Slowloris 类型的数据虽是少数类样本,但由于它们的流量特性,攻击流量特征明显,即 LOIC-UDP 为 UDP 类型流量且重复简单,Slowloris 为连接时间长且数据包单位时间内发送缓慢,除 SMOTE 方法外其他方法识别精确率、召回率和 F1 分数差距不大,故这两类不能作为参考。Bot 的流量特性是短时间内的重复流量,分析其特征会发现,它是交替发送两条流量,一条流量如同用户操作一般,而另一条流量是无效操作流量,在正常数据集和 SMOTE 过采样数据集中,由于重复发送的仿正常流量如正常流量一般,又没有明显的特征,故容易被识别为正常流量,容易漏报,而在 GAN 和 WDGP 过采样数据集中,生成器生成的流量会生成在仿正常流量周围,帮助模型更好地识别 Bot 类型流量。Web 和 XXS 的流量特性是大量且请求参数包含特殊字符,分析其特征可以发现,Web 和 XXS 与 Bot 相似,只不过 Web 是交替发送三条流量,两条无效和一条有效流量,而 XSS 也是交替发送两条流量。故与 Bot 相同,模型在 GAN 和 WDGP 过采样数据集上的综合入侵检测性能要比正常数据集和 SMOTE 过采样数据集更好。综合分析表 9~表 12 可得,WDGP 过采样方法在 1DCSP 多分类入侵检测和多分类少数类样本入侵检测实验结果中均表现出较好的过采样性能,可以有效提升物联网环境下轻量化入侵检测模型在多分类入侵检测中的性能。

3 结束语

针对物联网环境中边缘节点资源受限计算能力较低且模型性能普遍较差等特点,提出一种基于 WDGP 和 1DCSP 的入侵检测模型 WDGP-1DCSP。该模型首先结合了 WGAN-GP 和 DCGAN 的优点,通过 WGAN-GP 的损失函数和梯度惩罚机制与 DCGAN 的网络结构,使模型生成的数据质量显著提升,解决数据不平衡问题,为之后的轻量级多分类入侵检测模型训练扩展了数据集。然后,该模型在 1DCNN 基础上采用 CSP 框架,通过跨阶段局部融合特征方法,在降低模型复杂度和计算成本的同时,进一步提高了模型检测的准确率。最后,通过在多个数据集上进行的对比实验证明,所提模型 WDGP-1DCSP 能够在有限的计算和存储资源下实现高效的入侵检测,为未来物联网环境中的入侵检测研究提供借鉴。

虽然,所提模型 WDGP-1DCSP 在给定数据集上取得了不错的效果,但仍有许多方面今后应加以改进。在未来研究中,将考虑部署在真实的环境下进行测试,并引入更多轻量化技术以进一步降低模型的计算成本,并提高模型的泛化能力与检测入侵能力。

参考文献:

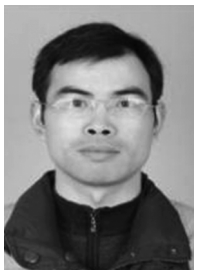
[1] Chen Z Y,Liu J X,Shen Y,et al. Machine Learning-Enabled Iot Security:Open Issues and Challenges under Advanced Persistent Threats[J]. ACM Computing Surveys,2023,55(5):105.

[2] Sarhan M,Layeghy S,Moustafa N,et al. Feature Extraction for Machine Learning-Based Intrusion Detection in IoT Networks[J]. Digital Communications and Networks,2024,10(1):205-216.

[3] Hadem P,Saikia D K,Moulik S. An SDN-Based Intrusion Detection System Using SVM with Selective Logging for IP

Traceback[J]. Computer Networks,2021,191:108015.
[4] 周杰英,贺鹏飞,邱荣发,等. 融合随机森林和梯度提升树的入侵检测研究[J]. 软件学报,2021,32(10):3254-3265.
[5] Li Y X,Zuo Y,Song H B,et al. Deep Learning in Security of Internet of Things [J]. IEEE Internet of Things Journal,2022,9(22):22133-22146.
[6] Xiao Y, Xing C, Zhang T, et al. An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks[J]. IEEE Access,2019,7:42210-42219.
[7] Zhao Z M,Li Z X,Jiang J L,et al. ERNN:Error-Resilient RNN for Encrypted Traffic Detection Towards Network-Induced Phenomena[J]. IEEE Transactions on Dependable and Secure Computing,2023,19(1):10036003.
[8] Khan I A, Moustafa N, Pi D, et al. A New Explainable Deep Learning Framework for Cyber Threat Discovery in Industrial IoT Networks [J]. IEEE Internet of Things Journal,2022,9(13):11604-11613.
[9] 项睿涵,潘巨龙,李玲艺,等. 一种物联网入侵检测和成员推理攻击研究[J]. 传感技术学报,2024,37(2):317-325.
[10] 孙红哲,王坚,王鹏,等. 基于 Attention-BiTCN 的网络入侵检测方法[J]. 信息网络安全,2024,24(2):309-318.
[11] Zhang H P,Huang L L,Wu C Q,et al. An Effective Convolutional Neural Network Based on Smote and Gaussian Mixture Model for Intrusion Detection in Imbalanced Dataset[J]. Computer Networks,2020,177:107315.
[12] Freitas de Araujo-Filho P,Kaddoum G,Campelo D R,et al. Intrusion Detection for Cyber-Physical Systems Using Generative Adversarial Networks in Fog Environment[J]. IEEE Internet of Things Journal,2021,8(8):6247-6256.
[13] Booij T M,Chiscop I,Meeuwissen E,et al. Ton_Iot:The Role of Heterogeneity and the Need for Standardization of Features and Attack Types in Iot Network Intrusion Data Sets[J]. IEEE Internet of Things Journal,2022,9(1):485-496.
[14] Kilincer I F, Ertam F, Sengur A. Machine Learning Methods for Cyber Security Intrusion Detection:Datasets and Comparative Study[J]. Computer Networks, 2021, 188:107840.
[15] Goodfellow I,Pouget-Abadie J,Mirza M,et al. Generative Adversarial Nets[C]//28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada,2014:2672-2680.
[16] Arjovsky M,Chintala S,Bottou L. Wasserstein Generative Adversarial Networks[C]//2017 34th International Conference on Machine Learning(ICML),Sydney,NSW,Australia,2017:214-223.
[17] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved Training of Wasserstein GANs[C]//31st International Conference on Neural Information Processing Systems, Long Beach,CA,USA,2017:5769-5779.
[18] Radford A,Metz L,Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks [C]//2016 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2016:arxiv.1511.06434.
[19] Wang C Y,Liao H Y M,Wu Y H,et al. CSPNet:A New Backbone That Can Enhance Learning Capability of CNN [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 2020:1571-1580.
[20] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV),Seoul,South Korea,IEEE,2019:1314-1324.
[21] 麻文刚,张亚东,郭进. 基于 LSTM 与改进残差网络优化的异常流量检测方法[J]. 通信学报,2021,42(5):23-40.
[22] Zhao J,Gui G,Xue Z,et al. A Novel Intrusion Detection Method Based on Lightweight Neural Network for Internet of Things[J]. IEEE Internet of Things Journal,2022,9(12):9960-9972.

卢志成(2000—),男,硕士研究生,研究方向为深度学习、入侵检测,812648135@qq.com;



徐海峰(1976—),男,硕士,副教授,研究方向为电子信息控制、物联网技术,3382462@qq.com;

潘巨龙(1965—),通讯作者。男,博士,教授,研究方向为物联网、智能计算和机器学习,pjl@cjlu.edu.cn。