一种基于深度可分离卷积和注意力机制的入侵检测方法

张志飞[1,2,3]，刘峰[1,2,3]，葛祎阳[1,2,3]，李烁[1,2,3]，张煜[4]，熊轲[1,2,3]

（1. 北京交通大学计算机与信息技术学院高速铁路网络管理教育部工程研究中心，北京 100044；
2. 北京交通大学轨道交通安全协同创新中心，北京 100044；3. 北京交通大学移动专用网络国家工程研究中心，北京 100044；
4. 国网能源研究院有限公司，北京 102209）

摘 要：为提高网络入侵检测中多分类的准确率，提出了一种基于深度可分离卷积和注意力机制的入侵检测方法。该方法通过深度可分离卷积、长短期记忆网络组成级联结构，提高了模型对数据中空间和时间特征的提取能力；进一步融入混合域注意力机制完善特征提取过程，提高了模型的检测能力。为了解决在中小样本上检测率低的问题，设计了一种基于变分自编码器和生成对抗网络的数据平衡策略，能有效应对网络数据集的数据不平衡现象，提升了所提检测方法的适应性。在 CICIDS-2017、NSL-KDD 和 UNSW-NB15 数据集上的实验结果表明，所提方法能够取得 99.80%、99.32%、83.87%的准确率，检测准确率分别提高了 0.6%、0.5%、2.3%。

关键词：深度学习；入侵检测；注意力机制；生成对抗网络

中图分类号：TN915.08

文献标志码：A

doi: 10.11959/j.issn.2096−3750.2023.00307

An intrusion detection method based on depthwise separable convolution and attention mechanism

ZHANG Zhifei[1,2,3], LIU Feng[1,2,3], GE Yiyang[1,2,3], LI Shuo[1,2,3], ZHANG Yu[4], XIONG Ke[1,2,3]

1. Engineering Research Center of Network Management Technology for High Speed Railway of Ministry of Education, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
2. Collaborative Innovation Center of Railway Traffic Safety, Beijing Jiaotong University, Beijing 100044, China
3. National Engineering Research Center of Advanced Network Technologies, Beijing Jiaotong University, Beijing 100044, China
4. State Grid Energy Research Institute Co., Ltd., Beijing 102209, China

Abstract: In order to improve the accuracy of multi-classification in network intrusion detection, an intrusion detection method was proposed based on depthwise separable convolution and attention mechanism. By constructing a cascade structure combining depthwise separable convolution and long-term and short-term memory networks, the spatial and temporal features of network traffic data can be better extracted. A mixed-domain attention mechanism was introduced to enhance the detection performance. To solve the problem of low detection rate in some samples, a data balance strategy based on the combination of the variational auto-encoder (VAE) the generative adversarial network (GAN) and was designed, which can effectively cope with imbalanced datasets and improve the adaptability of the proposed detection method. The experimental results show that the proposed method is able to achieve 99.80%, 99.32%, and 83.87% accuracy on the CICIDS-2017, NSL-KDD and UNSW-NB15 datasets, which is improved by 0.6%, 0.5%, and 2.3%, respectively.

Key words: deep learning, intrusion detection, attention mechanism, generative adversarial network

---

收稿日期：2022–08–04；**修回日期：**2022–10–29

通信作者：熊轲，kxiong@bjtu.edu.cn

基金项目：中央高校基本科研业务费资助项目（No.2022JBZY021）；国家自然科学基金资助项目（No.62071033）

Foundation Items: The Fundamental Research Funds for the Central Universities (No.2022JBZY021), The National Natural Science Foundation of China (No.62071033)

0 引言

随着计算机技术的日益普及，互联网上的用户数量和流量都在不断增长。在给用户带来数字化便利服务的同时，互联网面临的安全威胁也日益严峻。防火墙是一种已被广泛部署的常用网络安全技术手段，在一定程度上解决了网络中存在的安全访问控制问题，但仍存在诸多缺陷，如难以有效防御新型攻击、无法预防未备案的漏洞、自身易遭受攻击等。

为弥补防火墙技术的不足，入侵检测得到广泛关注。入侵检测能保护目标网络和计算机系统，监控被保护的系统中是否存在可疑行为或正在遭受攻击，并能够对入侵威胁做出主动反应措施。入侵检测可分为基于误用的入侵检测和基于异常的入侵检测，其中基于异常的入侵检测是较为主流的研究方向[1]。近年来，随着人工智能技术的不断发展，许多智能算法被应用于入侵检测领域，如机器学习算法等。Grammatikis 等[2]基于决策树构建入侵检测模型，在 CICIDS-2017 数据集上达到了 96.665%的准确率，并且将误报率降低至 1.145%。任晓奎等[3]采用粒子群算法优化了加权朴素贝叶斯算法，通过粗糙集理论降低数据维度，有效地提升了检测能力。

尽管如此，随着异常流量和数据的巨量化和高维化，基于传统机器学习的入侵检测方法已难以继续保持良好的检测性能。由于深度学习技术在处理海量、高维数据上的突出优势，越来越多的研究开始应用深度学习技术来提升入侵检测效果。Martin 等[4]设计了基于条件变分自编码器的深度入侵检测模型，将数据标签集成在解码器中，降低了模型的复杂性，提高了检测率。Wang 等[5]提出了一种名为 HAST-IDS 的深度入侵检测模型，结合卷积神经网络和长短期记忆网络在原始网络流文件上提取流量的空间和时序特征，无须使用额外的预处理技术。Althubiti 等[6]提出了一种基于长短期记忆网络的深度入侵检测模型，在 CICIDS-001 数据集上的测试结果表明该模型预测准确率能够达到 84.83%。Kanna 等[7]利用狮群算法优化了深度入侵检测模型的训练过程，引入了长短期记忆网络提取流量中的时间特征，在 NSL-KDD 数据集上达到了 90.67%的准确率。Khan 等[8]结合自编码器与长短期记忆网络设计了深度入侵检测模型，在 ISCX-UNB 数据集上的检测准确率达到了 97.52%。Jia 等[9]提出了将过采样与深度置信网络相结合的深度入侵检测模型，通过过采样算法增加数据集中小样本的数量，应用以深度置信网络作为主体的训练网络，并且引入了信息熵设置模型中的超参数，在 KDD99 数据集的训练和测试结果显示模型达到了 97.95%的准确率。

上述调研表明，深度学习在入侵检测领域已得到了研究者的广泛关注。然而，仍有诸多问题亟待研究与完善，如现有方法在小样本类别上的检测率仍然较差、对未知攻击检测效果不明、具体类别划分效果差等。

为此，提出了一种基于深度可分离卷积（DSC, depthwise separable convolution）和注意力机制的深度入侵检测方法，构建深度可分离卷积和长短期记忆网络相结合的级联结构初步提取特征，利用混合注意力机制进一步对特征进行处理，增强了模型的检测能力。此外，为了应对数据不平衡现象，提出了基于变分自编码器（VAE, variational auto encoder）和生成对抗网络（GAN, generative adversarial network）的数据平衡策略，以残差网络为模型，添加新的损失函数指导模型训练，有效地平衡了数据集，进一步提升了模型的准确率。所提深度入侵检测方法在 CICIDS-2017、NSL-KDD 和 UNSW-NB15 数据集上与其他基线方法进行了对比实验，验证了所提深度入侵检测方法的检测性能和在不同数据集上的适应性。

1 基于深度可分离卷积与注意力机制的深度入侵检测方法

本文设计了深度入侵检测方法，包含基于深度可分离卷积与注意力机制的深度入侵模型，以及基于变分自编码器和生成对抗网络的数据平衡策略。

1.1 基于深度可分离卷积与注意力机制的深度入侵模型

所设计的入侵检测模型包含数据预处理层、深度可分离卷积与双向长短期记忆（BiLSTM, bi-directional long short-term memory）网络的级联结构、混合域注意力机制层等结构。模型首先对网络流量数据进行预处理，得到适合模型输入的数据形式；接着将其输入到深度可分离卷积与长短期记忆网络的级联层中，充分提取数据的特征信息；随后通过混合域注意力机制，对提取的特征更深层次地学习；最后通过最大池化层、全连接层、Softmax 层。所提深度入侵检测模型的整体结构如图 1 所示，各结构的细节如下。

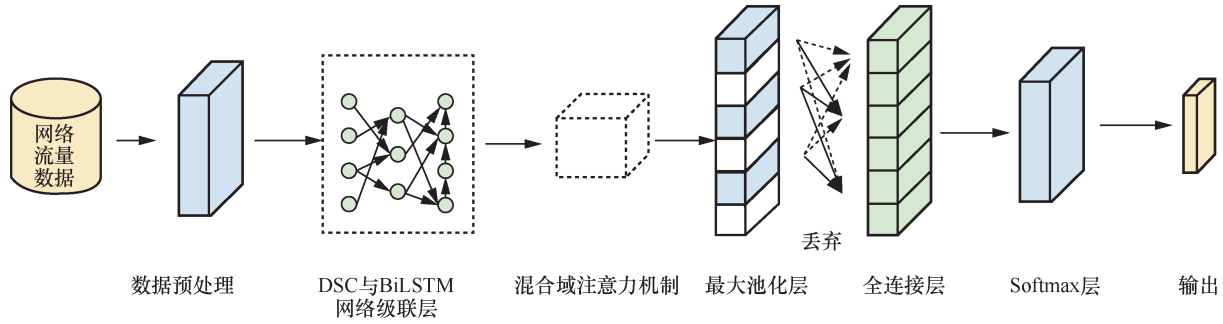


图 1　所提深度入侵检测模型的整体结构

1.1.1　深度可分离卷积与长短期记忆网络的级联结构

所采用的深度可分离卷积与长短期记忆网络的级联结构如图 2 所示，级联结构的细节如下。

1) 级联结构将输入的特征图分为 3 个分支，第 1 分支为多层深度可分离卷积，第 2 分支为多层长短期记忆网络，第 3 分支为深度可分离卷积和长短期记忆网络的块级串联结构。

2) 第 1 分支和第 2 分支的前两层分别为深度可分离卷积的串联结构和长短期记忆网络的串联结构。在第 1 分支和第 2 分支的第 3 层中，输入都为 2 个分支第 2 层的融合结果。第 4 层融合 2 个分支第 3 层的输出，形成包含时空信息的特征图。

3) 第三分支为深度可分离卷积和长短期记忆网络的块级串联结构，即 2 种网络串联作为一个整体，以整体为单位堆叠 2 次。

4) 在整个级联结构的第 5 层中，3 个分支特征合并成最终的特征图。在训练过程中，为应对协方差漂移现象，在每层卷积后进行批量归一化。

1.1.2　混合域注意力机制层

所设计的混合域注意力机制模型由空间注意力和通道注意力两个模块组成，如图 3 所示。通

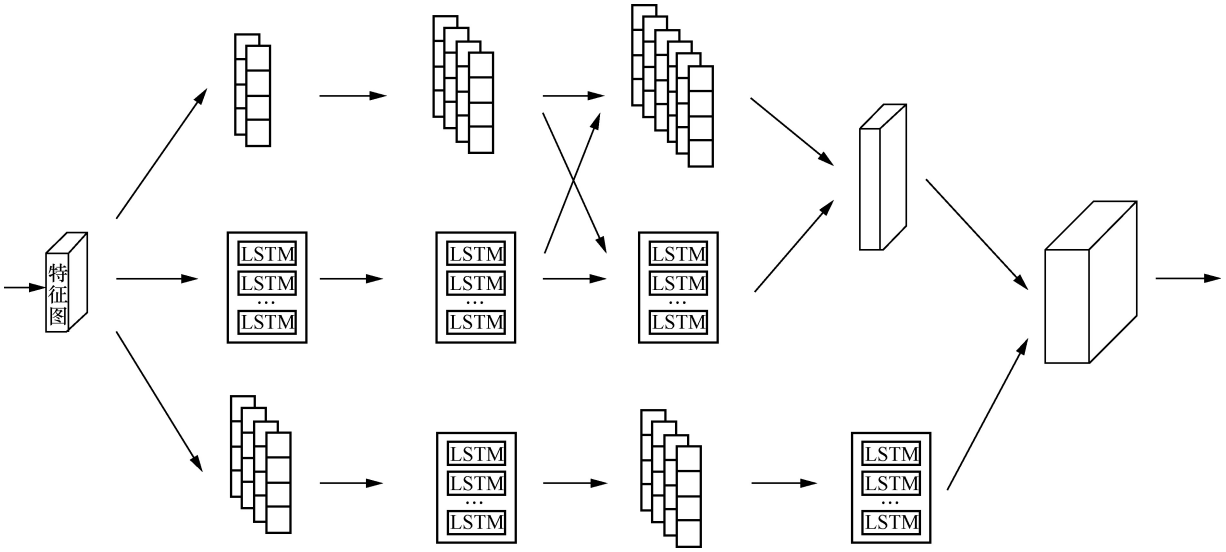


图 2　深度可分离卷积与长短期记忆网络的级联结构

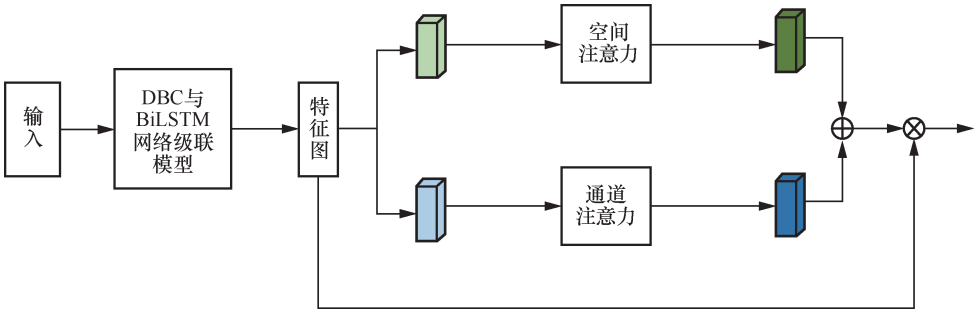


图 3　混合域注意力机制模型

过空间注意力和通道注意力两个模块对输入的特征图沿不同维度学习，得到相应的权重系数，将两个模块的计算结果叠加，得到权重矩阵，最后将权重矩阵和输入的特征向量相乘，得到最终的特征图。

本文采用的空间注意力模块结构如图 4 所示，该模块由池化操作、特征多方式合并、多尺度卷积组成，其细节如下。

1) 池化操作包含最大池化和平均池化，池化后得到两个 $H \times W$ 的特征图，其中 H 为特征图高度，W 为特征图宽度。

2) 特征多方式合并包含特征融合和特征相加两个分支。特征融合是在通道的方向上进行合并，得到 $H \times W \times 2$ 的特征空间关系图；特征相加是将两个特征图的信息进行相加，得到 $H \times W \times 1$ 的特征空间关系图。

3) 在特征融合后，结果分别输入到 3 个卷积核大小分别为 3×3、4×4、5×5 的多尺度卷积层，然后将 3 种特征结果相加。计算过程如下。

$$\mathrm{SA}_1 = \sigma\left(K^{3\times3}\left(\left[F^{\mathrm{s}}_{\mathrm{avg}}; F^{\mathrm{s}}_{\mathrm{max}}\right]\right)\right) \tag{1}$$

$$\mathrm{SA}_2 = \sigma\left(K^{4\times4}\left(\left[F^{\mathrm{s}}_{\mathrm{avg}}; F^{\mathrm{s}}_{\mathrm{max}}\right]\right)\right) \tag{2}$$

$$\mathrm{SA}_3 = \sigma\left(K^{5\times5}\left(\left[F^{\mathrm{s}}_{\mathrm{avg}}; F^{\mathrm{s}}_{\mathrm{max}}\right]\right)\right) \tag{3}$$

$$\mathrm{SA}_c = \left(\mathrm{SA}_1 \oplus \mathrm{SA}_2 \oplus \mathrm{SA}_3\right) \tag{4}$$

其中，SA_1、SA_2、SA_3 分别表示多尺度卷积结果，$F^{\mathrm{s}}_{\mathrm{avg}}$、$F^{\mathrm{s}}_{\mathrm{max}}$ 分别表示通道方向上的平均值和最大值，$\oplus$ 表示计算特征图的相加，$K^{3\times3}$、$K^{4\times4}$ 和 $K^{5\times5}$ 分别表示卷积核的大小，σ 表示激活函数 sigmoid，SA_c 表示特征融合后进行多尺度卷积的最终特征图。

4) 合并两个分支的结果，得到最终的空间注意力系数，计算过程为

$$\mathrm{SA} = \sigma\left(\mathrm{SA}_c \oplus \mathrm{SA}_a\right) \tag{5}$$

其中，SA 表示特征相加后经过卷积操作的特征图。

本文采用的通道注意力模块结构如图 5 所示。通道注意力机制针对输入特征中的通道关系计算关系系数。所采用的通道注意力模块分为池化操

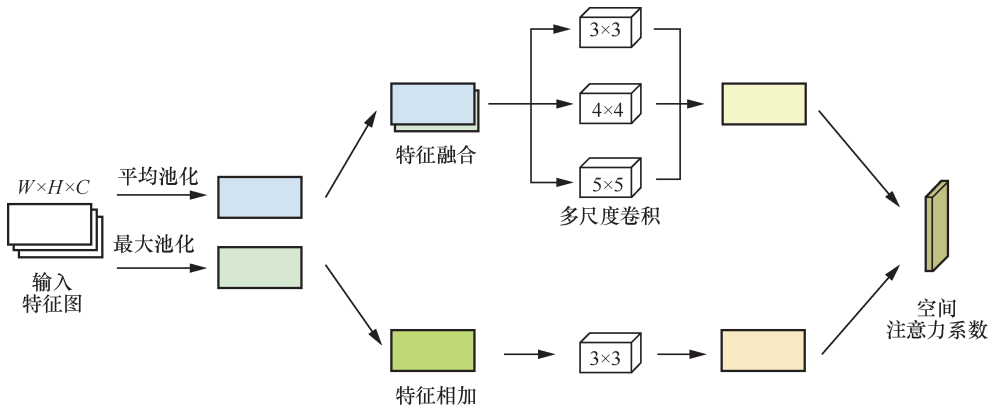


图 4　空间注意力模块结构

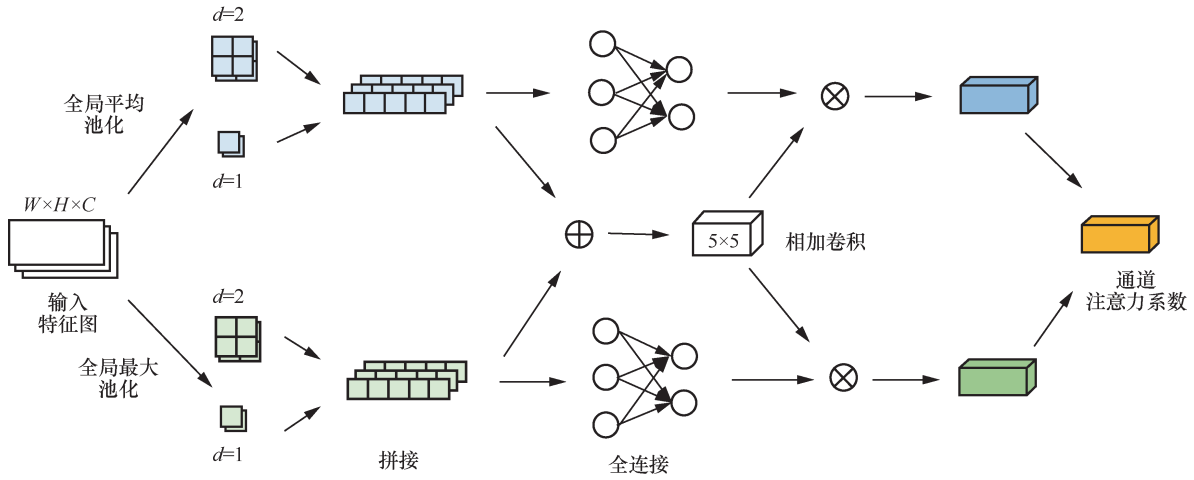


图 5　通道注意力模块结构

作、拼接、全连接、相加卷积 4 个部分。

1) 池化操作分为全局平均池化和全局最大池化，每个池化操作由多个尺度组成，其中所设置的尺度参数为 1 和 2 两种。

2) 池化后对特征进行拼接，形成 5 维特征向量，之后通过全连接层，拟合通道间的关联关系；同时进行相加卷积操作，对复合特征进行深层次提炼。最后将全连接后的特征向量与相加卷积的结果相乘。

3) 计算得到两个不同层面的通道注意力系数，相加得到最终的结果。获取通道注意力系数过程中涉及的计算为

$$\mathrm{CA}_1=\sigma\left(W_{\mathrm{avg}}^{\mathrm{c}}\cdot F_{\mathrm{avg}}^{\mathrm{c}}+b_{\mathrm{avg}}^{\mathrm{c}}\right) \tag{6}$$

$$\mathrm{CA}_2=\sigma\left(W_{\mathrm{max}}^{\mathrm{c}}\cdot F_{\mathrm{max}}^{\mathrm{c}}+b_{\mathrm{max}}^{\mathrm{c}}\right) \tag{7}$$

$$\mathrm{CA}_{\mathrm{c}}=\sigma\left(K^{5\times5}\left(\left[F_{\mathrm{avg}}^{\mathrm{c}};F_{\mathrm{max}}^{\mathrm{c}}\right]\right)\right) \tag{8}$$

$$\mathrm{CA}_{\mathrm{avg}}=\mathrm{CA}_1\otimes\mathrm{CA}_{\mathrm{c}} \tag{9}$$

$$\mathrm{CA}_{\mathrm{max}}=\mathrm{CA}_2\otimes\mathrm{CA}_{\mathrm{c}} \tag{10}$$

$$\mathrm{CA}=\mathrm{CA}_{\mathrm{avg}}\oplus\mathrm{CA}_{\mathrm{max}} \tag{11}$$

其中，$F_{\mathrm{avg}}^{\mathrm{c}}$、$F_{\mathrm{max}}^{\mathrm{c}}$ 分别表示不同池化上不同尺度操作的拼接结果，$W_{\mathrm{avg}}^{\mathrm{c}}$、$W_{\mathrm{max}}^{\mathrm{c}}$ 分别表示全连接层的权重系数，$b_{\mathrm{avg}}^{\mathrm{c}}$、$b_{\mathrm{max}}^{\mathrm{c}}$ 分别表示全连接层的偏置，CA_1、CA_2 分别表示池化后全连接操作的计算结果，σ 表示激活函数 sigmoid，CA_{c} 表示相加卷积操作，$\otimes$ 代表矩阵乘法操作，$\mathrm{CA}_{\mathrm{avg}}$、$\mathrm{CA}_{\mathrm{max}}$ 分别表示全连接层的结果与相加卷积相乘结果，CA 表示最终的通道注意力系数。

1.2　基于变分自编码器和生成对抗网络的数据平衡策略

为解决数据集中数据不平衡的问题，设计了基于变分自编码器和生成对抗网络数据的平衡策略，其模型结构如图 6 所示。

模型分为生成器、判别器和分类器 3 个部分。生成器读取真实数据及对应标签，生成伪造数据；判别器判断生成器输出数据的真伪；分类器输出数据的类别。具体模型细节如下。

1) 生成器

生成器基于变分自编码器设计，包含编码器和解码器，都由残差网络构成。在编码器中，残差网络提取数据中隐含的特征，将其转化到潜在表征 Z 中。解码器的输入为潜在表征 Z 与类别标签，将类别标签转换为独热编码，拼接到潜在表征 Z 尾部，生成特定类别数据。

2) 判别器和分类器

判别器与分类器的结构基本一致，都由残差网络构成。

3) Loss 结构图

基于变分自编码器和生成对抗网络的数据平衡模型 Loss 结构如图 7 所示，各组件采用的 Loss 函数细节如下。

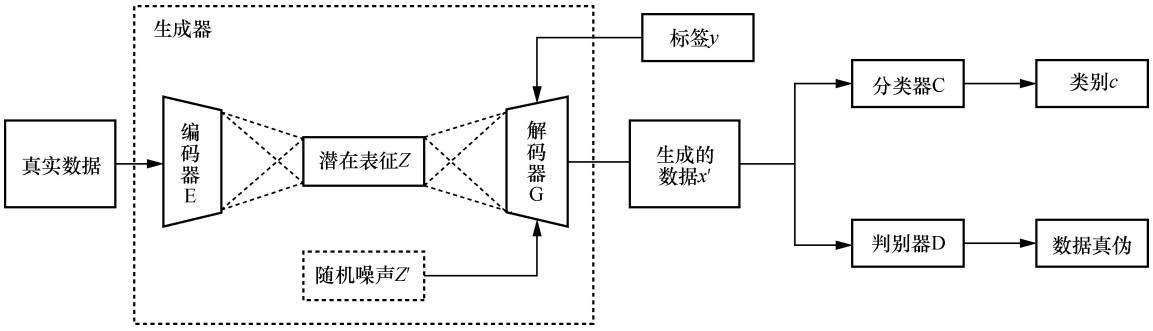


图 6　基于变分自编码器和生成对抗网络的数据平衡模型

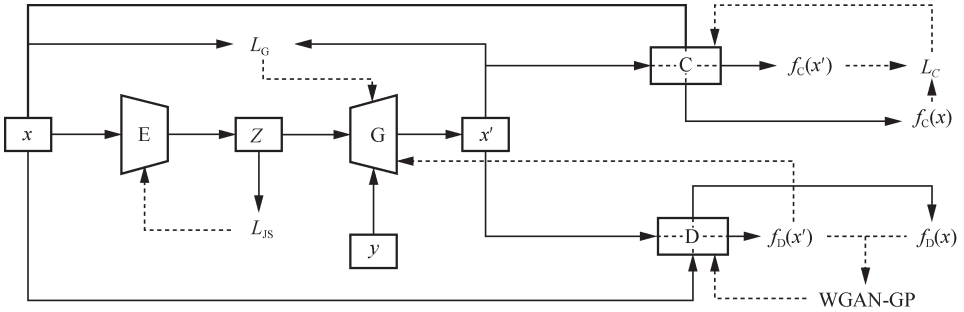


图 7　基于变分自编码器和生成对抗网络的数据平衡模型 Loss 结构

分类器的 Loss 函数 L_{c} 如式(12)所示，通过将真实数据和重构数据分别传入模型，最小化差值进行优化。

$$L_{\mathrm{c}}=1/2N\sum_{n=1}^{N}\left\|f_{\mathrm{c}}\left(x_n\right)-f_{\mathrm{c}}\left(x_n'\right)\right\|_2^2 \tag{12}$$

其中，N 为样本的数量，x_n 表示真实样本，x_n' 表示生成样本，$f_{\mathrm{c}}(\bullet)$ 表示分类器的输出。

判别器的 Loss 函数 L_{D} 如式(13)所示，采用 Wassertein 距离解决梯度不稳定问题，并进一步约束重构数据与输入数据的差异，优化判别器的网络。

$$L_{\mathrm{D}}=\mathbb{E}_{x\sim P_r}D(x)-\mathbb{E}_{z'\sim P_{z'}}D\left(G\left(z'\right)\right) \tag{13}$$

其中，$x\sim P_r$ 表示真实样本 x 服从从真实样本分步 P_r，$z'\sim P_{z'}$ 表示输入噪声 z' 服从噪声分步 $P_{z'}$，$D(\bullet)$ 表示判别器的输出，$G(\bullet)$ 表示生成器的输出。

生成器的 Loss 函数 L_{G} 如式(14)所示。L_{G} 分为两部分，第一部分采用平方损失减小重构数据和真实数据之间的差异值，第二部分为伪造数据的分数。

$$L_{\mathrm{G}}=1/N\sum_{n=1}^{N}\left\|\left(x_n-x_n'\right)\right\|_2^2-\mathbb{E}_{z'\sim P_{z'}}D\left(G\left(z'\right)\right) \tag{14}$$

编码器的 Loss 函数 L_{JS} 为 JS 散度为

$$\begin{aligned}L_{\mathrm{JS}}=&1/2\mathrm{KL}\left(q\left(z|\ x_r\right)\|\left(q\left(z|\ x_r\right)+p_z\right)/2\right)+\\&1/2\mathrm{KL}\left(p_z\|\left(q\left(z|\ x_r\right)+p_z\right)/2\right)\end{aligned} \tag{15}$$

其中，z 表示真实数据的潜在表征，$q\left(z|\ x_r\right)$ 表示输入为 x_r 时潜在表征 z 的概率，p_z 表示潜在表征 z 的概率，$\mathrm{KL}(\bullet)$ 表示 KL 散度。

基于变分自编码器和生成对抗网络的数据平衡模型训练流程见算法 1。

算法 1 基于变分自编码器和生成对抗网络的数据平衡模型训练流程

初始化 判别器 D 和分类器 C 的参数，编码器 E 和解码器 G 的参数。

while 生成器未收敛或者没有到达预设的迭代次数时 do

for $m=1,\cdots,M$

for $n=1,\cdots,3$

从真实数据中采样 $x\sim P_x$；

用 x 训练 C 网络，更新 C 的参数；

生成假数据 $\hat{x}$ 和随机数 α，将假数据和真实数据制成混合数据，$x'=\alpha\hat{x}+(1-\alpha)x$；

训练 D 网络，采用式(13)中的损失函数更新 D 的参数。

end for

真实数据中采样 $x\sim P_x$，从噪声数据中采样 $z\sim P_z$，采用 MSE 和 JS 散度更新 E 和 G 网络；

以判别器和分类器网络来优化 E 和 G 的参数；

采用 Adam 算法，优化 C 的参数。

end for

end while

完成训练之后，将数据平衡模型应用到数据平衡策略中平衡数据集。所提数据平衡策略流程如图 8 所示。

数据集的平衡分为 4 个步骤。首先选出需要扩充的数据类别，计算出需要扩充的数据量，接着利用 SMOTE 算法进行数据扩充，随后将扩充后的数据和原数据合并，最后利用数据平衡模型扩充剩余的数据。

以 UNSW-NB15 数据集为例，UNSW-NB15 平衡效果见表 1。

表 1 UNSW-NB15 平衡效果

类别	数据平衡前占比	总生成数量/个	数据平衡后占比
Fuzzers	6.0%	3 000	7.0%
Worms	0.4%	3 000	1.4%
Shellcode	3.6%	4 200	5.0%
Generic	9.0%	5 400	10.8%

本文选取 UNSW-NB15 中具有代表性的 4 种小样本类别进行展示，这 4 类数据占比提升在 1.0%～1.8%，其中 Worms 类别占比由 0.4%提升至 1.4%，Shellcode 类别由 3.6%提升至 5.0%，说明本文所提数据平衡模型能够有效地平衡数据集。

2 实验及分析

2.1 数据集及预处理

本文的实验数据集采用 NSL-KDD[10]、CICIDS-2017[11] 以及 UNSW-NB15[12]。其中，

图 8 所提数据平衡策略流程

NSL-KDD 数据集共包含 148 516 条流量数据，CICIDS-2017 数据集共包含 2 830 743 条流量数据，UNSW-NB15 数据集包含 257 673 条流量数据。3 种数据集的格式各不相同，因此需要进行不同的处理。其中，NSL-KDD 数据集已经过预处理，可直接使用，CICIDS-2017 数据集和 UNSW-NB15 数据集都需要预处理，步骤如下。

1) 数据清洗

数据集中存在个别数据丢失的现象，即某些属性没有内容，或被填充为空格、NaN 以及 Infinity 等。为保证训练效果，本文使用均值填充非法格式数据。

2) 字符数值化与标签数字化

在 UNSW-NB15 中，存在一些特征为字符型的数据，如“proto”属性，内容为“tcp”“udp”“icmp”，将这些字符标签转换为 1、2、3。类似地，将其他数据集类别标签转换为相应的数字标记。

3) 数据归一化

数据集中存在多个属性，每个属性中的数值大小及范围有差异。若直接将未归一化的数据送入模型学习，易导致值域范围大的数据特征具有很高的权重，使其成为主导属性，而值域范围小的数据的权重小，易使得特征丢失。为此，本文使用数据归一化算法对数据集进行完善，计算式为

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{16}$$

其中，x'表示归一化之后的值，x 表示初始特征值，$x_{\min}$ 表示该属性中的最小特征值，$x_{\max}$ 表示该属性中的最大特征值。

由于数据集中属性较多，本文以 UNSW-NB15 数据集为例，选取了其中具有代表性的 6 个属性，UNSW-NB15 数据集预处理效果见表 2。其中类别及属性 1 经标签数字化转换为数字形式，属性 2、属性 3 及属性 6 值域范围相对较大，经数据归一化后有效降低数值大小，避免其在后续训练中成为模型主导属性而影响模型效果。

2.2 实验分析

本文实验使用的评价指标包含准确率、召回率及 F1 值。

准确率表示所有预测正确（正类负类）的样本占所有样本数的百分比，反应模型的整体预测能力，计算式为

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{17}$$

召回率表示正确预测为正的样本占全部实际为正的百分比，也被称为查全率，代表正例样本中的覆盖率，计算式为

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

F1 值是关于召回率和精确率的函数，即两者的调和平均数，能够综合考虑整体数据与具体样本，计算式为

$$\text{F1} = 2 \times \frac{\text{Precision}\ \ \text{Recall}}{\text{Precision}\ +\ \text{Recall}} \tag{19}$$

其中，精确率为 $\text{Precision} = \dfrac{\text{TP}}{\text{TP} + \text{FP}}$。

2.2.1 所提深度入侵检测模型在 3 个数据集上的实验分析

添加混合域注意力机制前后的整体准确率对比见表 3。

表 3　添加混合域注意力机制前后的整体准确率对比

模型	CICIDS-2017	NSL-KDD	UNSW-NB15
所提入侵检测模型	**99.12%**	**98.86%**	**81.07%**
无注意力机制模型	98.78%	98.14%	79.26%

可以看出，所提入侵检测模型在 3 个数据集上的检测率分别为 99.12%、98.86%、81.07%，检测

表 2　UNSW-NB15 数据集预处理效果

阶段	类别	属性 1 state	属性 2 sttl	属性 3 Dtcpb	属性 4 synack	属性 5 ackdat	属性 6 dmean
预处理前	Normal	FIN	62	17 824 254	0.071 3	0.069 7	123
	Reconnaissance	INT	254	0	0	0	0
	Generic	FIN	62	2.43×10^9	0.022 7	0.480	1 133
预处理后	0	4	0.243	0.004 15	0.022 1	0.023 8	0.082
	1	2	0.996	0	0.005 47	0.028 5	0.029 3
	5	4	0.243	0.565	0.007 05	0.016 4	0.756

效果较好。较未添加混合域注意力机制的模型相比，检测准确率有所提升。

本文所提出入侵检测模型与其他带有注意力机制的基线模型的对比结果见表 4～表 6。

所提模型与其他带有注意力机制的基线模型在 CICIDS-2017 数据集上的准确率对比见表 4。在 CICIDS-2017 数据集上，各模型的准确率达到 97%以上，其中，所提模型能够取得较优的成绩，达到 99.12%。

所提模型与其他带有注意力机制的基线模型在 NSL-KDD 数据集上的准确率对比见表 5。在 NSL-KDD 数据集上，所提模型的检测率最高，达到了 98.86%，最差的 LCHI+AM 模型准确率不足 80%。实验结果表明所提模型在 NSL-KDD 数据集上性能较优。

所提模型与其他带有注意力机制的基线模型在 UNSW-NB15 数据集上的准确率对比见表 6，可以看出，在 UNSW-NB15 数据集上，各模型的检测效果都不足 85%，所提模型总体准确率排在第二位。

表 4 所提模型与其他带有注意力机制的基线模型在 CICIDS-2017 数据集上的准确率对比

模型	准确率
LCHI+AM[13]	97.97%
CN+AM[14]	97.56%
L2+AMNN[15]	99.05%
FL+AM[16]	99.28%
DQN+AM[17]	98.70%
本文模型	**99.12%**

表 5 所提模型与其他带有注意力机制的基线模型在 NSL-KDD 数据集上的准确率对比

模型	准确率
LCHI+AM[13]	79.16%
CN+AM[14]	95.88%
DQN+AM[17]	97.40%
DLNID[18]	90.73%
DL+AM[19]	95.00%
本文模型	**98.86%**

表 6 所提模型与其他带有注意力机制的基线模型在 UNSW-NB15 数据集上的准确率对比

模型	准确率
LCHI+AM[13]	80.33%
TCN+AM[20]	72.92%
DAL[21]	81.28%
DL+AM[19]	73.00%
CAL+AM[22]	78.94%
本文模型	**81.07%**

2.2.2 数据平衡后的实验对比分析

从上述实验可以看出，在一些数据集（如 UNSW-NB15 数据集）上，所有模型都未能取得较好的检测性能，这可能是 UNSW-NB15 数据集中存在数据不平衡现象所致。为此，将所提模型分别在平衡前后的数据集上进行训练，所提模型在平衡数据集前后的整体准确率对比如图 9 所示。

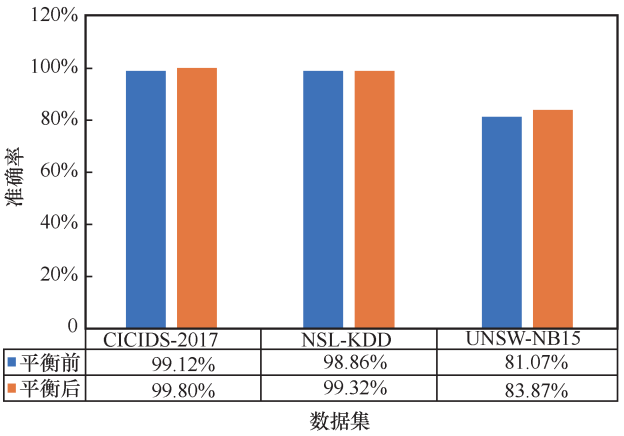


图 9 所提模型在平衡数据集前后的整体准确率对比

从图 9 可以看出，当采用所提数据平衡策略后，整体准确率较未做数据平衡均有所提升，表明所提出的数据平衡策略能够有效平衡数据集，进而提高整体检测性能。

为了进一步测试所提入侵检测方法的检测性能，本文还对所提方法与其他基线方法在 3 种数据集上的检测性能进行对比。其中，本文模型与其他模型在 CICIDS-2017 数据集上的 F1 值对比见表 7，本文模型与其他模型在 NSL-KDD 数据集上的 F1 值对比见表 8，本文模型与其他模型在 UNSW-NB15 数据集上的召回率对比见表 9。

表 7 展示了在 CICIDS-2017 数据集上所提入侵检测方法与 Hierarchical[23]、WISARD[24]、Forest

PA[25]、LIBSVM[26]以及 FURIA[27]5 种方法的 F1 分数对比。所提入侵检测方法在 14 个攻击类型中有 9 个具有最高的 F1 分数。值得注意的是，在一些类别上，基线方法存在检出率为 0 的情况，而所提方法能够较好地检出各个类别，表明在 CICIDS-2017 数据集上，所提入侵检测方法效果较优。

表 8 展示了所提入侵检测方法与其他方法在 NSL-KDD 数据集上的 F1 值对比，其中对比的 5 种方法分别为 DL[28]、RNN-ADV[29]、GAN-PSO-ELM[30]、DCNN[31]、SAVAER-DNN[32]。在 4 个攻击类别中，本文所提方法在 U2R、R2L 以及 Dos 上的效果最好，而在 Normal 中排在第二位，所提入侵检测方法在 NSL-KDD 数据集上检测效果较优。

表 9 为所提入侵检测方法与其他方法在 UNSW-NB15 数据集上的召回率对比，3 种对比方法分别为 MDPCA-DBN[33]、SE[34]以及 GTDR[35]。可以看出所提入侵检测方法在 9 个攻击类别中有 5 个效果最好，表明所提入侵检测方法在 UNSW-NB15 数据集上具有较强的检测能力。

表 7　　本文模型与其他模型在 CICIDS-2017 数据集上的 F1 值对比

标签	本文方法	Hierarchical	WISARD	ForestPA	LIBSVM	FURIA
BENIGN	**99.85%**	98.86%	97.13%	96.45%	94.87%	96.83%
DoS Hulk	**99.99%**	96.78%	67.60%	93.94%	73.70%	98.65%
Port Scan	99.78%	**99.88%**	51.40%	99.59%	48.52%	87.11%
DdoS	**99.98%**	99.87%	54.69%	99.81%	55.97%	99.75%
DoS GoldenEye	**99.65%**	67.57%	48.71%	67.57%	57.57%	65.14%
FTP-Patator	**99.93%**	99.63%	0.00%	99.72%	0.00%	99.63%
SSH-Patator	99.84%	99.90%	0.00%	**100%**	0.00%	**100%**
DoS slowloris	**99.30%**	97.75%	78.90%	92.84%	78.18%	93.75%
DoS Slowhttptest	**99.34%**	93.84%	23.35%	86.82%	76.56%	78.35%
Bot	**92.82%**	46.47%	1.44%	48.71%	0.00%	48.07%
Web Attack Brute Force	**82.14%**	73.26%	4.69%	73.46%	80.81%	49.79%
Web Attack-XSS	50.00%	30.62%	1.25%	34.37%	0.00%	**58.75%**
Infiltration	76.92%	**100%**	50.00%	83.33%	0.00%	83.33%
Web Attack Sql Injection	**100%**	50.00%	0.00%	50.00%	0.00%	50.00%
Heartbleed	80.00%	**100%**	80.00%	**100%**	0.00%	40.00%

表 8　　本文模型与其他模型在 NSL-KDD 数据集上的 F1 值对比

标签	本文方法	DL	RNN-ADV	GAN-PSO-ELM	DCNN	SAVAER-DNN
Normal	98.82%	98.11%	95.46%	97.85%	**99.47%**	95.30%
Dos	**99.78%**	98.75%	96.61%	98.11%	99.13%	85.10%
Probe	**98.65%**	83.34%	85.55%	97.31%	94.35%	74.47%
R2L	**94.10%**	48.35%	55.58%	89.28%	83.21%	53.59%
U2R	**84.00%**	74.28%	63.92%	80.53%	64.10%	44.50%

表 9 本文模型与其他模型在 UNSW-NB15 数据集上的召回率对比

标签	本文方法	MDPCA-DBN	SE	GTDR
Normal	94.65%	82.85%	91.82%	**97.39%**
Generic	**98.70%**	96.93%	98.32%	81.37%
Exploits	**90.71%**	83.51%	85.00%	76.22%
Fuzzers	60.15%	44.39%	60.97%	**64.42%**
DoS	10.50%	23.72%	**25.00%**	14.29%
Reconnaissance	**79.00%**	76.68%	74.80%	46.07%
Analysis	8.76%	0.00%	11.00%	**20.45%**
Backdoor	9.44%	0.85%	10.79%	**67.32%**
Shellcode	**69.53%**	39.47%	58.22%	36.39%
Worms	**40.00%**	11.11%	37.50%	18.37%

3 结束语

将深度可分离卷积结构和长短期记忆网络结构级联应用到网络入侵检测，通过构建深度可分离卷积和长短期记忆网络相结合的级联结构，提取流量数据的时间和空间特征，提升了入侵检测的准确率。

引入混合域注意力机制进一步提高了模型在各类别中的检出率。所设计的混合域注意力机制包含通道和空间两个模块。通道注意力机制由多尺度卷积和相加卷积组成，空间注意力机制由特征多方式合并和多尺度卷积组成，充分挖掘潜在信息。所提模型较未引入注意力机制的模型整体准确率有提升。

采用基于变分自编码器和生成对抗网络的数据平衡策略平衡数据集，提升了对少数类样本的检出率。在传统生成对抗网络模型基础上，基于变分自编码器设计生成器，引入残差网络，并设计新的损失函数，更有效地平衡数据集。采用所提的数据平衡策略后，所提入侵检测方法在 3 种数据集上检测的整体准确率提升到了 99.80%、99.32%以及 83.87%，并且少数类的检出率也有提升，提高了所提方法在不同数据集上的适应性。

参考文献：

[1] LIU H Y, LANG B. Machine learning and deep learning methods for intrusion detection systems: a survey[J]. Applied Sciences. 2019, 9(20): 4396-4420.

[2] RADOGLOU-GRAMMATIKIS P I, SARIGIANNIDIS P G. An anomaly-based intrusion detection system for the smart grid based on CART decision tree[C]//Proceedings of 2018 Global Information Infrastructure and Networking Symposium (GIIS). Piscataway: IEEE Press, 2018: 1-5.

[3] 任晓奎, 缴文斌, 周丹. 基于粒子群的加权朴素贝叶斯入侵检测模型[J]. 计算机工程与应用, 2016, 52(7): 122-126.
REN X K, JIAO W B, ZHOU D. Intrusion detection model of weighted navie Bayes based on particle swarm optimization algorithm[J]. Computer Engineering and Applications, 2016, 52(7): 122-126.

[4] LOPEZ-MARTIN M, CARRO B, SANCHEZ-ESGUEVILLAS A, et al. Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT[J]. Sensors (Basel, Switzerland), 2017, 17(9): E1967.

[5] WANG W, SHENG Y Q, WANG J L, et al. HAST-IDS: learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection[J]. IEEE Access, 2018, 6(99): 1792-1806.

[6] ALTHUBITI S A, JONES E M, ROY K. LSTM for anomaly-based network intrusion detection[C]//Proceedings of 2018 28th International Telecommunication Networks and Applications Conference (ITNAC). Piscataway: IEEE Press, 2018: 1-3.

[7] KANNA P R, SANTHI P. Unified deep learning approach for efficient intrusion detection system using integrated spatial-temporal features[J]. Knowledge-Based Systems, 2021, 226: 107132.

[8] ASHFAQ KHAN M, KIM Y. Deep learning-based hybrid intelligent intrusion detection system[J]. Computers, Materials & Continua, 2021, 68(1): 671-687.

[9] JIA H P, LIU J, ZHANG M, et al. Network intrusion detection based on IE-DBN model[J]. Computer Communications, 2021, 178: 131-140.

[10] SIFRE L, MALLAT S. Rigid-Motion scattering for texture classification[J]. Computer Science, 2014, 3559: 501-515.

[11] MNIH V, HEESS N, GRAVES A. Recurrent models of visual attention[C]//Advances in neural information processing systems, 2014(2): 2203-2212.

[12] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Communications of the ACM, 2020, 63(11): 139-144.

[13] LEI S W, XIA C H, WANG T B. LCHI: low-order correlation and high-order interaction integrated model oriented to network intrusion detection[J]. Wireless Communications and Mobile Computing, 2021, 2021: 6830372.

[14] 刘烁, 张兴兰. 基于双重注意力的入侵检测系统[J]. 信息网络安全, 2022, 22(1): 80-86.
LIU S, ZHANG X L. Intrusion detection system based on dual attention[J]. Netinfo Security, 2022, 22(1): 80-86.

[15] 曹磊, 李占斌, 杨永胜, 等. 基于双层注意力神经网络的入侵检测方法[J]. 计算机工程与应用, 2021, 57(19): 142-149.
CAO L, LI Z B, YANG Y S, et al. Intrusion detection method based on two-layer attention networks[J]. Computer Engineering and Applications, 2021, 57(19): 142-149.

[16] CHEN Z, LV N, LIU P F, et al. Intrusion detection for wireless edge networks based on federated learning[J]. IEEE Access, 2020(8): 217463-217472.

[17] SETHI K, MADHAV Y V, KUMAR R, et al. Attention based multi-agent intrusion detection systems using reinforcement learning[J]. Journal of Information Security and Applications, 2021, 61: 102923.

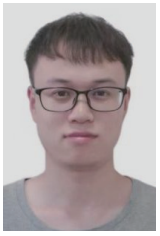
[18] FU Y F, DU Y S, CAO Z J, et al. A deep learning model for network intrusion detection with imbalanced data[J]. Electronics, 2022, 11(6): 898.

[19] ANDRESINI G, APPICE A, CAFORIO F P, et al. ROULETTE: a neural attention multi-output model for explainable Network Intrusion Detection[J]. Expert Systems With Applications, 2022, 201: 117144.

[20] ZHAO P, FAN Z J, CAO Z W, et al. Intrusion detection model using temporal convolutional network blend into attention mechanism[J]. International Journal of Information Security and Privacy, 2022, 16(1): 1-20.
[21] CAO K, ZHU J Q, FENG W, et al. Network intrusion detection based on dense dilated convolutions and attention mechanism[C]//Proceedings of 2021 International Wireless Communications and Mobile Computing (IWCMC). Piscataway: IEEE Press, 2021: 463-468.
[22] 曹轲, 朱金奇, 马春梅, 等. 联合多重卷积与注意力机制的网络入侵检测[J]. 天津师范大学学报(自然科学版), 2021, 41(3): 75-80.
CAO K, ZHU J Q, MA C M, et al. Network intrusion detection based on multiple convolutions and attention mechanism[J]. Journal of Tianjin Normal University (Natural Science Edition), 2021, 41(3): 75-80.
[23] AHMIM A, MAGLARAS L, FERRAG M A, et al. A novel hierarchical intrusion detection system based on decision tree and rules-based models[C]//Proceedings of 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS). Piscataway: IEEE Press, 2019: 228-233.
[24] DE GREGORIO M, GIORDANO M. An experimental evaluation of weightless neural networks for multi-class classification[J]. Applied Soft Computing, 2018, 72: 338-354.
[25] ADNAN M N, ISLAM M Z. Forest PA: constructing a decision forest by penalizing attributes used in previous trees[J]. Expert Systems With Applications, 2017, 89: 389-403.
[26] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2007: 2-20.
[27] ZHANG X Q, CHEN J H, ZHOU Y, et al. A multiple-layer representation learning model for network-based attack detection[J]. IEEE Access, 2019(7): 91992-92008.
[28] MOHAMMADI S, NAMADCHIAN A. A new deep learning approach for anomaly base IDS using memetic classifier[J]. International Journal of Computers Communications & Control, 2017, 12(5): 677.
[29] QURESHI A U H, LARIJANI H, YOUSEFI M, et al. An adversarial approach for intrusion detection systems using Jacobian saliency map attacks (JSMA) algorithm[J]. Computers, 2020, 9(3): 58.
[30] SUMAIYA THASEEN I, ASWANI KUMAR C. Intrusion detection model using fusion of Chi-square feature selection and multi class SVM[J]. Journal of King Saud University - Computer and Information Sciences, 2017, 29(4): 462-472.
[31] 丁红卫, 万良, 周康, 等. 基于深度卷积神经网络的入侵检测研究[J]. 计算机科学, 2019(10): 173-179.
DING H W, WAN L, ZHOU K, et al. Study on intrusion detection based on deep convolution neural network[J]. Computer Science, 2019(10): 173-179.
[32] ZHANG G L, WANG X D, LI R, et al. Network intrusion detection based on conditional Wasserstein generative adversarial network and cost-sensitive stacked autoencoder[J]. IEEE Access, 2020(8): 190431-190447.
[33] YANG Y Q, ZHENG K F, WU C H, et al. Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks[J]. Applied Sciences, 2019, 9(2): 238.
[34] RAJAGOPAL S, KUNDAPUR P P, HAREESHA K S. A stacking ensemble for network intrusion detection using heterogeneous datasets[J]. Security and Communication Networks, 2020, 2020: 4586875.
[35] PAPAMARTZIVANOS D, GÓMEZ MÁRMOL F, KAMBOURAKIS G. Dendron: genetic trees driven rule induction for network intrusion detection systems[J]. Future Generation Computer Systems, 2018, 79: 558-574.

[作者简介]

张志飞（1971– ），男，博士，北京交通大学计算机与信息技术学院高级工程师，主要研究方向为无线通信、网络安全等。



刘峰（1998– ），男，北京交通大学计算机与信息技术学院硕士生，主要研究方向为网络安全、入侵检测、深度学习等。

葛祎阳（1997– ），男，北京交通大学计算机与信息技术学院博士生，主要研究方向为深度学习、强化学习、信息年龄、无线能量传输网络和 6G 网络等。

李烁（1998– ），男，北京交通大学计算机与信息技术学院硕士生，主要研究方向为可靠传输、拥塞控制、强化学习等。

张煜（1983– ），男，博士，国网能源研究院有限公司高级研究员，主要研究方向为边缘计算、无线协作网络和能源互联网等。

熊轲（1981– ），男，博士，北京交通大学计算机与信息技术学院教授、副院长，主要研究方向为无线协作网络、无线移动网络和网络信息理论等。