研究与开发

一种适用于小样本条件的网络入侵检测方法

胡炜晨，许聪源，詹勇，陈广辉，刘思情，王志强，王晓琳
（嘉兴学院信息科学与工程学院，浙江 嘉兴 314001）

摘　要：现有的网络入侵检测技术多数需要大量恶意样本用于模型训练，但在现网实战时，往往只能获取少量的入侵流量样本，属于小样本条件。对此，提出了一种适用于小样本条件的网络入侵检测方法。该方法由数据包采样模块和元学习模块两部分组成，数据包采样模块用于对网络原始数据进行筛选、剪切与重组，元学习模块则用于特征提取、结果分类。在基于真实网络流量数据源构建的 3 个小样本数据集上的实验结果表明，该方法适用性好、收敛快，能有效减少异常点的出现，在 10 个训练样本下的检测率最高可达 99.29%，准确率最高可达 97.93%，相比目前已有的算法，分别提升了 0.12%和 0.37%。

关键词：入侵检测；小样本；元学习；网络安全；深度学习
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2023166

A network intrusion detection method designed for few-shot scenarios

HU Weichen, XU Congyuan, ZHAN Yong, CHEN Guanghui, LIU Siqing, WANG Zhiqiang, WANG Xiaolin
College of Information Science and Engineering, Jiaxing University, Jiaxing 314001, China

Abstract: Existing intrusion detection techniques often require numerous malicious samples for model training. However, in real-world scenarios, only a small number of intrusion traffic samples can be obtained, which belong to few-shot scenarios. To address this challenge, a network intrusion detection method designed for few-shot scenarios was proposed. The method comprised two main parts: a packet sampling module and a meta-learning module. The packet sampling module was used for filtering, segmenting, and recombining raw network data, while the meta-learning module was used for feature extraction and result classification. Experimental results based on three few-shot datasets constructed from real network traffic data sources show that the method exhibits good applicability and fast convergence and effectively reduces the occurrence of outliers. In the case of 10 training samples, the maximum achievable detection rate is 99.29%, while the accuracy rate can reach a maximum of 97.93%. These findings demonstrate a noticeable improvement of 0.12% and 0.37% respectively, in comparison to existing algorithms.

Key words: intrusion detection, few-shot, meta-learning, network security, deep learning

收稿日期：2023−04−11；**修回日期**：2023−08−21
通信作者：许聪源，cyxu@zjxu.edu.cn
基金项目：浙江省自然科学基金资助项目（No.LQ23F020006，No.LQ22F020004）
Foundation Items: The Natural Science Foundation of Zhejiang Province (No.LQ23F020006, No.LQ22F020004)

0 引言

网络入侵手段层出不穷，极大地威胁了网络安全。网络入侵检测是实现网络安全的重要内容之一，也是保障网络安全的重要手段。在当下的网络环境中，网络入侵检测是颇具复杂性与挑战性的工作。各种机器学习算法，特别是深度学习算法，被认为可以用于检测大规模网络流量中的入侵行为[1]。

大部分研究人员建立的入侵检测模型在KDD99、CICIDS2017、ISCX2012 等被广泛使用的入侵检测数据集上都可以达到较高的检测率。现有的机器学习算法在面对大量样本时，有非常出色的表现，但是在实际的网络环境中，并没有这样理想。Zhang 等[2]认为基于深度学习的流量检测方法依赖于大量的样本数据，Li 等[3]认为深度网络从少量数据中学习新概念的能力有限。在面对新出现的入侵时，能获取的样本数量往往非常少，无法获取足够多且已标记的样本来制作数据集，使得现有的基于深度学习的入侵检测算法很难发挥原有的作用。而且对每一种新的入侵手段制作相应的数据集是非常困难的，一方面是时间上的限制，制作这样一个数据集需要大量的时间；另一方面是人力物力的限制，需要大量的资源才能制作一个数据集。因此，已广泛使用的基于深度学习的检测方法都难以在小样本条件下有效解决入侵检测技术存在的问题，而能对网络安全构成威胁的往往就是这种无法获取足够样本的新的入侵方式。若要维护网络安全，如何实现小样本条件下的网络入侵的有效检测是亟待解决的问题。

小样本条件下的机器学习算法在图像分类领域已经有了一些进展，例如，Zhang 等[4]提出的不确定性感知小样本图像分类方法，利用数据独立的不确定性建模，来降低噪声对小样本学习的不良影响；Afrasiyabi 等[5]引入基于混合的特征空间学习，以在小样本图像分类的背景下获得丰富而稳健的特征表示；Kang 等[6]提出的关系嵌入网络结合自关联和交叉关联两个关系模块，学习端到端管理器中的关系嵌入。但是在入侵检测领域，针对小样本场景的研究还非常有限。因此，本文针对性地提出了一种小样本入侵检测方法，可以利用它来检测只有少量样本的入侵行为。

本文的主要贡献如下。

- 提出了一种适用于小样本条件的网络入侵检测方法。在 10 个训练样本下的检测率最高可达 99.29%，准确率最高可达 97.93%，优于现有其他方法。
- 实验使用 ISCX2012 、CICIDS2017 、CICIDS2018 数据集的原始流量，并提出了将原始流量转换成小样本数据集的方法，转换后的数据适用于评估小样本检测方法。

1 相关工作

1.1 网络入侵检测

近年来，网络入侵检测已经成为网络研究领域的一个热点。传统的网络入侵检测方法分为两种：基于规则的检测方法和基于负载特征的检测方法。这些网络入侵检测技术存在一些问题，如缺乏灵活性、误报或漏报率较高等。

随着机器学习方法的不断进步，与之相关的方法也被引进入侵检测领域中。例如，Aldwairi 等[7]在网络入侵检测技术中结合了一种被称为受限玻尔兹曼机的机器学习技术；Abdelmoumin 等[8]提出了一种优化技术来增强使用单学习器的基于异常的机器学习入侵检测系统的性能。基于机器学习的检测技术的一大局限是需要设计一个能准确反映数据特征的特征集。这个特征集的质量对整个模型的性能有着决定性的影响，但设计一个好的特征集是十分困难的。

然而，深度学习的出现，克服了传统机器学习面临的困难，深度学习能够使模型自动化地学习到有效的特征。深度学习的方法自推出以来就

在众多研究领域上取得了巨大的成功。越来越多的研究人员将深度学习引入网络入侵检测技术，Haghighat 等[9]提出一种新型基于投票的深度学习的框架，可以利用任何类型的深度学习结构，并且提供了聚合最佳模型的能力。Basati 等[10]提出使用一种轻量级和高效的基于深度特征提取思想的神经网络，在该模型中，网络的输入向量被排列在 3D 空间中，其各个值彼此靠近，可以搭建更轻量化的模型结构。Soltani 等[11]提出一种深度入侵检测系统，在该系统的学习和检测阶段使用了流量的元数据和纯文本，使系统可以挖掘到在流量中被自动提取的特征之间的复杂关系。

1.2　小样本网络入侵检测

当出现一种新的网络入侵手段时，如上所述，大部分检测模型并不能准确并且快速地识别出这些入侵。这个难题被定义为在小样本条件下的网络入侵检测问题。

近几年，有学者提出一些有关小样本学习的算法。这些小样本学习算法着重解决深度神经网络依赖大样本的问题，并且吸引更多的学者去研究小样本问题。如上所述，随着小样本学习研究的兴起，已经有学者在网络入侵检测技术中采用小样本学习方法。例如，Liang 等[12]提出一种优化的基于类内/类间的变分小样本学习模型，其中，基于变分贝叶斯来优化类内距离的近似值，基于特征融合的相似最大化用于优化类间距离；Xu 等[13]提出一种基于元学习框架的检测方法，设计了一个由特征提取网络和比较网络组成的深度神经网络；Iliyasu 等[14]提出一种利用有监督的自动编码器判别表示学习的方法；Yang 等[15]提出一个多任务表示增强元学习模型，将监督学习和基于聚类的无监督学习结合，以提升来自少量标记数据的加密流量表示的差异性；Ouyang 等[16]基于原型网络提出了一种新的小样本学习入侵检测算法，通过一种协调独热编码和主成分分析的新方法来预处理数据集；Yu 等[17]提出一种基于分层数据的卷积神经网络，第一层从原始流量中自动提取抽象特征，第二层从数据包进一步构建表示；Zhang 等[2]通过利用协方差矩阵表征每个流量类别，并根据协方差度量函数计算查询流量与每个类别之间的相似度来实现小样本入侵检测； Wang 等[18]提出一种新的孪生网络，设计的深度学习网络能捕捉流量特征的动态关系；Gamal 等[19]提出一种基于小样本深度学习的入侵检测系统，可自动识别来自网络边缘的零日攻击；Shi 等[20]提出一种基于模型无关的元学习（model-agnostic meta-learning，MAML）的入侵检测框架，从原始流量中提取统计和序列特征，并引入学习遗忘衰减机制来动态控制冲突的影响；Ye 等[21]提出一种用于语义感知流量检测的小样本潜在狄利克雷生成学习的方法，使用基于潜在狄力克雷分配的伪样本生成算法增强训练数据，并提出一种模糊回收方法，提高基于长短期记忆（long short-term memory，LSTM）的分类器的鲁棒性；Verkerken 等[22]提出一种分层入侵检测多阶段方法；Xu 等[23]提出一种基于度量的一阶元学习框架，通过多个任务训练入侵检测模型，以最大化模型的泛化能力。

本文在上述研究的基础上，进一步降低检测需要的样本数量，提出的方法使用了更少的训练样本。此外，为了更贴合实战环境，考虑在不同网络下进行实验，即使用不同网络下采集的数据集进行实验和测试，如使用 CICIDS2017 数据集进行训练，并使用 ISCX2012 数据集进行测试。

2　小样本网络入侵检测方法

本文所提出的小样本网络入侵检测方法，将元学习的思想引入网络入侵检测领域。首先，根据官方文档标注网络原始数据，以 16×16×16 的比例采样数据流，并重构为数组，随后将处理好的数据划分为支持集和查询集。在训练时，模块 N 中的参数 θ 根据支持集中的数据进行第一次更新，并根据第一次更新与查询集中的数

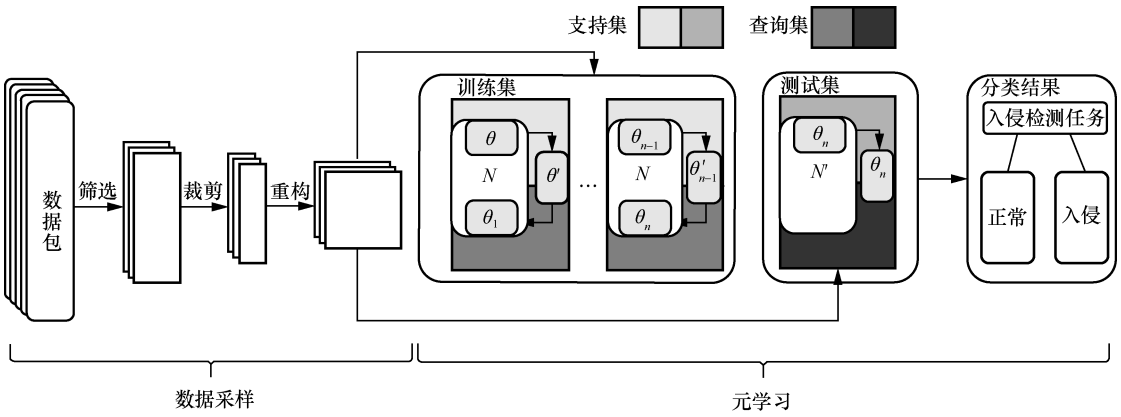


图 1　本文方法流程框架

据进行第二次更新并反馈到模块 N 中。在测试阶段，复制训练完成的模块 N 得到模块 N'，对于模块 N' 中的参数 θ_n，只根据支持集中的数据进行一次更新，查询集中的数据仅用作评估。同时测试阶段能对测试集中的任务输出模型的分类结果，分类结果为正常和入侵两类。本文方法流程框架如图 1 所示。

2.1　数据预处理

大多数研究工作利用了数据集中已经提取好特征的子集合，如 CICIDS2017 数据集，用 CICFlowMeter 软件提取了 80 多个特征。但是本文直接使用数据集提供的 pcap 包。pcap 包中包含了全部原始流量，可以使模型在训练过程中充分地学习到更多的有效特征。

首先，本文根据官方提供的可扩展标记语言（extensible markup language，XML）文件对 pcap 包中的数据打标签。根据 XML 文件中所提供的数据流的源 IP 地址、源端口、目的 IP 地址和目的端口，对每条数据流打上相应的标签。为了得到适用于小样本网络入侵检测的数据集，构建了 ISCX2012AS（after samping，数据采样后）数据集和 CICIDS2017AS 数据集。ISCX2012AS 数据集中包含了正常流量和 4 类攻击，其中攻击包括内部网络渗透、超文本传送协议（hypertext transfer protocol，HTTP）拒绝服务攻击、使用互联网中继交谈（Internet relay chat，IRC）僵尸网络的分布式拒绝服务（distributed denial of service，DDoS）攻击、暴力破解安全外壳（secure shell，SSH）；CICIDS2017AS 数据集中包含了正常流量和 5 类攻击，其中攻击包括暴力破解文件传送协议（file transfer protocol，FTP），暴力破解 SSH，使用 Slowloris、Slowhttptest、Hulk、GoldenEye 进行的拒绝服务（denial of service，DoS）攻击，端口扫描攻击，使用 LOIT 的 DDoS 攻击。

进一步地，对上述已经打过标签的数据进行裁剪，对每条数据流都提取前 16 个数据包，每个数据包提取前 256 byte，并且将提取的每条数据处理为 16×16×16 的数组。然后，为了得到相对平衡的数据集，将提取过的数据打乱，并且按照相同比例提取每一种类型的数据。

然后，使用归一化方法将偏差过大的数值变换到 0-1 分布内，转化后的数据集按照 8:2 的比例随机切割为训练集和测试集，再从训练集中取 20%作为查询集，剩余的作为支持集。

最后，为了更好地处理网络入侵检测中样本数量少的问题，引入了元学习任务的概念，将类型随机组合形成多个小任务，从而充分利用已有样本的信息，在随机过程中选取任意一天中的正常和入侵数据作为测试集，其余数据均作为训练集，任务分类如图 2 所示。

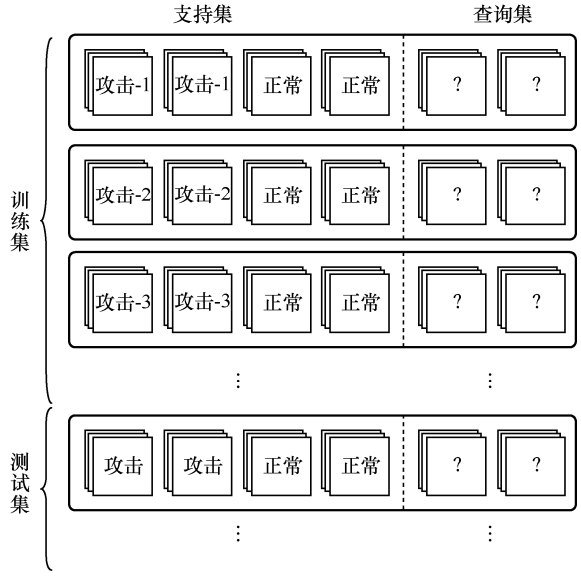


图 2　任务分类

2.2　元学习算法

在机器学习任务中，某些内部特征比其他内部特征更容易被学习，而相似的任务之间往往存在相同或者相似的特征。对于比较容易学习的特征，让模型先学习相似的任务，得到一个能快速学习具体任务的元模型，快速学习的元模型在样本数较少的具体任务上也有不错的学习能力。基于这种观点，设计如下算法：在训练阶段，支持集中的数据用于参数 θ 的第一次更新，得到新的参数 θ'，但不更新到模块 N 内部。第一次更新与查询集中的数据用于第二次更新，更新模块 N 内部的 θ 得到 θ_1，反复迭代得到最终参数 θ_n。在测试阶段，复制训练完成的模块 N 得到 N'。N' 中的参数只进行第一次更新，并在测试阶段输出对测试集中查询集的数据的分类结果，分类结果有正常和入侵两类。

对于第一次更新，传统的元学习算法只进行一次更新，本文采用了多次更新的思想，多次更新的结果再反馈到第二次更新过程中，具体如式（1）所示。其中，E 表示数学期望，$\tau \sim p(\tau)$ 表示入侵检测任务，α 表示学习率，$\circ$ 表示标量乘法，∇ 表示梯度算子，$\mathcal{L}_{\text{sup}(\tau)}(\theta)$ 表示训练时模型在支持集上的损失率，$\mathcal{L}_{\text{qry}(\tau)}(\theta')$ 表示训练后的模型在查询集上的损失率。

$$\min_{\theta,\alpha} E_{\tau\sim p(\tau)}[\mathcal{L}_{\text{qry}(\tau)}(\theta')] = E_{\tau\sim p(\tau)}[\mathcal{L}_{\text{qry}(\tau)}(\theta-\alpha\circ\nabla\mathcal{L}_{\text{sup}(\tau)}(\theta))] \tag{1}$$

详细的训练过程如算法 1 所示，首先从处理好的数据集中反复随机抽取入侵检测任务 T_i，组成一个 $T_i \sim T_n$ 的任务池作为元学习算法的训练集。本文的元学习算法基于二重梯度，在参数更新过程中可以调整两个学习率 α 与 β，以快速得到元模型。算法 1 的第二个 for 循环是第一次更新，更新得到参数 θ'。参数 θ' 将用于第二次更新过程。第一次更新出的参数并不会更新到模型内部，而第二次更新则是对模型内部参数 θ 的更新。在第一次更新中会有两次梯度更新。损失函数 $\mathcal{L}_{\mathcal{T}_i}$ 是交叉熵算法。更新 $\theta \leftarrow \theta - \beta\nabla_\theta \sum_{\mathcal{T}_i\sim\mathcal{T}_n} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$ 是第二次更新。第二次更新与第一次更新存在两处不同，一是第一次更新只用了一个任务的损失函数去梯度更新，而第二次更新利用第一次更新得到的所有损失函数结果的总和进行梯度更新，梯度更新采用的算法是 Adam 算法；二是在于参与计算的样本，为了加强模型的泛化能力，避免在支持集上过拟合，第二次更新用查询集的样本进行计算。

算法 1　训练算法

输入　入侵检测任务分布 $\mathcal{T}_n$，学习率 α、β，更新次数 update

随机初始化 θ

for all epoch do

入侵检测任务批次 $\mathcal{T}_i \sim \mathcal{T}_n$

for all $\mathcal{T}_n$ do

来自 $\mathcal{T}_n$ 的包含 K 个样本的样本集 $D=\{X^{(j)},Y^{(j)}\}$

根据 K 个样本和损失函数 $\mathcal{L}_{\mathcal{T}_i}$ 计算梯度 $\mathcal{L}_{\mathcal{T}_i}(f_\theta)$

计算梯度下降自适应函数：$\theta' = \theta - \alpha\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$

for j=1 to update do

根据 K 个样本和损失函数 $\mathcal{L}_{\mathcal{T}_i}$ 计算梯度 $\mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$

计算梯度下降的自适应参数：

$\theta' = \theta' - \alpha\nabla_{\theta'}\mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$

end

end

更新 $\theta \leftarrow \theta - \beta\nabla_{\theta}\sum_{\mathcal{T}_i\sim\mathcal{T}_n}\mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$

end

模型的训练过程如式 2 所示。

$$p\left(D_T^{\text{qry}}|\theta_0, D_T^{\text{trn}}\right) = \prod_{\tau\in T} p\left(D_\tau^{\text{qry}} \mid \theta'_\tau = \theta_0 + \alpha\nabla_{\theta_0}\ln p\left(D_\tau^{\text{sup}} \mid \theta_0\right)\right) \tag{2}$$

其中，$p\left(D_\tau^{\text{qry}}|\theta'_\tau\right) = \prod_{i=1}^{|D_T^{\text{qry}}|} p(y_i|x_i,\theta'_\tau)$，$p(y_i|x_i,\theta'_\tau)$ 表示模型内部参数为 θ'_τ 时，输入 x_i，模型输出为 y_i 的概率；D_T^{trn} 表示训练集；D_τ^{sup} 表示训练支持集；D_τ^{qry} 表示训练查询集。θ'_τ 通过支持集不断逼近模型的后验 $p(\theta_\tau|\theta_0, D_\tau^{\text{sup}})$，在用查询集验证模型时，$\theta'_\tau$ 又成为模型的先验。

面对新的任务，在元模型的基础上，测试模型的过程与训练的过程大致相同，不同的地方主要在于以下两点。

（1）在随机初始化 θ 时，不用再随机初始化参数，训练模型是直接训练完成的元模型的复制。初始参数即训练完成的元模型中的参数。

（2）测试过程需要跳过更新 $\theta \leftarrow \theta - \beta\nabla_{\theta}\sum_{\mathcal{T}_i\sim\mathcal{T}_n}\mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$，因为任务的查询集是用来测试模型的，标签对于模型是未知的。因此，测试过程没有第二次梯度更新，而是直接利用第一次梯度计算的结果更新参数。

3 实验与分析

3.1 数据集

本文使用两个公开数据集作为流量数据的来源，分别是 CICIDS2017[24]与 ISCX2012[25]。CICIDS2017 数据集由加拿大网络安全研究所（Canadian Institute for Cybersecurity，CIC）于 2017 年 7 月 3 日至 7 日采集，包含 5 天的数据流量，总共 2 830 473 个网络流量，除了正常的流量，还包含入侵的流量。网络数据的格式由原始网络数据包（pcap 包）和 CICFlowMeter 提取的数值统计特征组成，更加类似于真实流量。ISCX2012 数据集由新不伦瑞克大学 AliShiravi 等人在 2012 年创建，旨在为网络入侵检测构建当代基准。该数据集通过监测 7 天网络活动得来，由正常流量和恶意流量组成。该数据集能反映现实的网络和流量，并且具备多样化的入侵场景。生成该数据集的方法通过完全捕获网络跟踪，使得数据集的自然性得以保留。

目前，还缺乏小样本的基准数据集，所以利用这两个数据集的原始数据，通过本文方法中的数据采样，制作了 ISCX2012AS 和 CICIDS2017AS 两个小样本数据集。为了区分两个数据集，用小写字母（a、b、c 等）表示来自 ISCX2012 的数据，用大写字母（A、B、C 等）表示来自 CICIDS2017 的数据，ISCX2012AS 和 CICIDS2017AS 数据集中包含的攻击类型分别见表 1 和表 2。

表 1 ISCX2012AS 数据集中包含的攻击类型

名称	攻击类型
攻击-a	内部网络渗透
攻击-b	HTTP 拒绝服务攻击
攻击-c	使用 IRC 僵尸网络的 DDoS 攻击
攻击-d	暴力破解 SSH
正常	正常流量

表 2 CICIDS2017AS 数据集中包含的攻击类型

名称	攻击类型
攻击-A	暴力破解 FTP
攻击-B	暴力破解 SSH
攻击-C	使用 Slowloris、Slowhttptest、Hulk、GoldenEye 的 DoS 攻击
攻击-D	端口扫描攻击
攻击-E	使用 LOIT 的 DDoS 攻击
正常	正常流量

3.2　实验设置

本文实验中用到的软/硬件环境如下：CPU 为 Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60 GHz，内存为 128 GHz，操作系统为 Ubuntu 20.04，GPU 为 NVIDIA RTX3090，显存为 24 GB。采用 CUDA 11.3 作为 GPU 加速库，使用了 Python 3.8 和深度学习框架 PyTorch 1.1.0。

在本文方法中，构造模块 N 作为特征提取器与分类器。其中，特征提取器用于处理 16 通道 16×16 的数据。模块 N 的结构如图 3 所示。特征提取器的输出为 64×1×1，其中还使用批标准化（batch normalization，BN）约束数据正态化让数据分布更集中，最后使用线性整流函数（rectified linear unit，ReLU）作为激活函数。

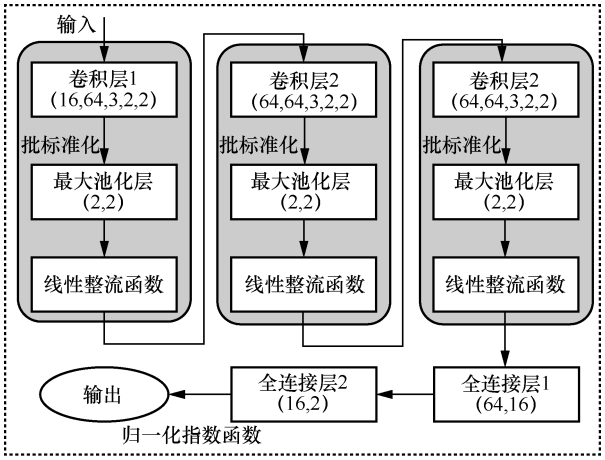


图 3　模块 N 的结构

其中，卷积层的参数分别是输入通道数、输出通道数、卷积核的尺寸、步长与填充，最大池化层的参数分别是窗口大小和步幅，全连接层的参数分别是输入的样本大小、输出的样本大小。

算法的超参数设置见表 3。

根据训练集与测试集的数据来源，设计了如下两类实验，实验Ⅰ为同网实验，表示在同种网络情况下，各种攻击类型和样本数量在两种数据集上的检测情况。而在真实场景中，所获取的数据往往来自不同的网络，即存在跨网情况，因此设计实验Ⅱ为跨网实验，通过改变训练集与测试集的网络来源，模拟真实场景下的小样本条件。具体实验如下。

表 3　算法的超参数设置

超参数	值
α	4×10^{-2}
β	2×10^{-4}
epoch	1 000
update	3
batch-size	8

实验Ⅰ：在两个数据集上分别进行实验，保证在实验Ⅰ中的训练集与测试集均来自同一个数据集，即用同一个网络环境中的数据检测相同网络环境情况下的攻击类型。ISCX2012AS 包含 4 种攻击类型。使用其中 3 种作为训练集的数据，另一种攻击类型作为测试集的数据。构建足够数量的训练和测试任务进行实验。相应地，CICIDS2017AS 数据集包含 5 种攻击类型。将其中 4 种作为本方法训练集的数据，剩下 1 种作为测试集的数据。分别提供 4 组和 5 组的平行实验，对所求得的数据取均值作为评判性能的标准。由于每个实验是相互独立的，所以可以在多台计算机上同时进行实验。

为了系统性探究小样本问题，考虑了典型的样本数量，设置了 K=5、10，此外，为了进一步探究极小样本的情况，还设置了 K=3 进行研究，分别在 ISCX2012AS 和 CICIDS2017AS 这两个数据集上进行实验。

实验Ⅱ：仅将 ISCX2012AS 或 CICIDS2017AS 中的一个作为训练集，另一个数据集作为测试集，保证训练集与测试集中的数据来自不同的网络，用训练集与测试集的来源不同来表示跨网，其余参数与实验Ⅰ中保持一致，最后与实验Ⅰ中的结果进行比较。

综上所述，面对来自不同网络的数据、不同的软/硬件环境、不同的攻击类型，实验Ⅱ更具有挑战性和实用性。

3.3 检测结果

实验Ⅰ：在 ISCX2012AS、CICIDS2017AS 数据集上的结果分别见表 4 和表 5，将准确率（ACC）与检测率（DR）作为衡量标准。可以看出，与传统的监督学习算法需要大量样本进行训练不同，本文方法在面对不同攻击类型时均有不错的表现力，并且在 K=10 的情况下，两种数据集上的平均检测率能分别达到 97.70%、98.13%。这说明了使用本文方法实现小样本网络入侵检测的可行性。

与此同时，可以得出以下两个结论。

（1）不同数据集的选择会对实验结果产生一定的影响。对于同一个数据集内部不同攻击类型的选取，笔者发现在 ISCX2012AS 中攻击-b 作为测试集的效果较差，在 CICIDS2017AS 中攻击-C 作为测试集的效果较差。对于不同数据集的选取，笔者发现当样本数量 K 为 3 或 5 时，在 ISXC2012AS 数据集上的表现均比 CICIDS2017AS 的表现好。随着样本数量的增加，当样本数量为 10 时，CICIDS2017AS 数据集上的平均检测率反而优于 ISXC2012AS。

（2）本文方法仅需少量样本就能达到较高的检测水准。对于这两种数据集来说，若 K 值不断增加，ACC 会不断提高。对于 ISCX2012AS、CICIDS2017AS 来说，在 K=3 时平均准确率就分别达到了 94.55%和 93.38%，平均检测率分别达到了 95.17%和 94.86%；当 K=10 时，ISCX2012AS 数据集上的平均准确率可达 98.53%，CICIDS2017AS 数据集上的平均检测率达 98.13%。

实验Ⅱ：通过改变数据集的类别，设置了跨网实验：在 ISCX2012AS 上进行训练，在 CICIDS2017AS 上进行测试，在 ISCX2012AS 上的跨网检测结果见表 6。在 CICIDS2017AS 上进行训练，在 ISXC2012AS 上进行测试，在 CICIDS2017AS 的跨网检测结果见表 7。

表 4 在 ISCX2012AS 数据集上的检测结果

名称	K=3		K=5		K=10	
	ACC	DR	ACC	DR	ACC	DR
攻击-a	97.58%	97.99%	99.96%	99.92%	99.30%	98.81%
攻击-b	88.32%	88.74%	99.27%	98.93%	99.06%	98.13%
攻击-c	93.10%	94.73%	95.22%	94.80%	97.70%	97.83%
攻击-d	99.19%	99.21%	97.25%	97.49%	98.06%	96.03%
平均值	94.55%	95.17%	97.93%	97.79%	98.53%	97.70%

表 5 在 CICIDS2017AS 数据集上的检测结果

名称	K=3		K=5		K=10	
	ACC	DR	ACC	DR	ACC	DR
攻击-A	94.14%	94.61%	98.60%	97.28%	99.60%	99.24%
攻击-B	98.54%	97.87%	99.17%	99.60%	99.66%	99.87%
攻击-C	80.60%	88.09%	84.23%	90.19%	92.14%	93.61%
攻击-D	97.40%	97.48%	97.50%	97.57%	99.99%	99.99%
攻击-E	96.20%	96.27%	97.75%	97.72%	98.24%	97.94%
平均值	93.38%	94.86%	95.45%	96.47%	97.93%	98.13%

表 6　在 ISCX2012AS 上的跨网检测结果

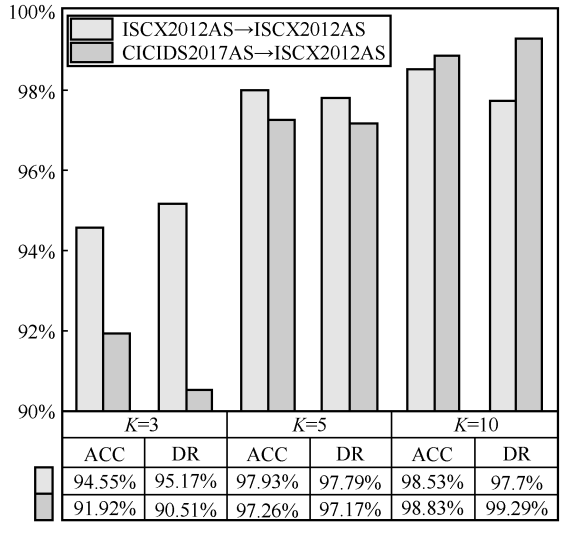
训练集→测试集	*K*=3		*K*=5		*K*=10	
	ACC	DR	ACC	DR	ACC	DR
A,B,C,D,E→a	99.99%	99.99%	99.99%	99.99%	99.95%	99.99%
A,B,C,D,E→b	86.23%	84.41%	95.07%	95.56%	98.90%	99.27%
A,B,C,D,E→c	96.14%	95.82%	99.85%	99.79%	97.50%	99.99%
A,B,C,D,E→d	85.30%	81.80%	94.14%	93.34%	98.96%	97.90%
平均值	91.92%	90.51%	97.26%	97.17%	98.83%	99.29%

表 7　在 CICIDS2017AS 上的跨网检测结果

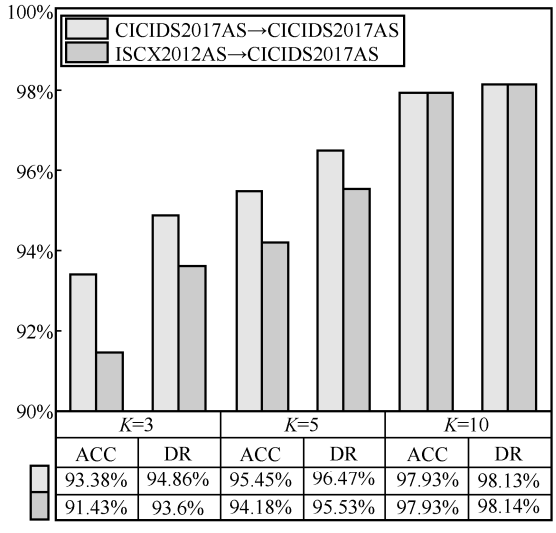
训练集→测试集	*K*=3		*K*=5		*K*=10	
	ACC	DR	ACC	DR	ACC	DR
a,b,c,d,→A	92.80%	92.69%	95.75%	98.17%	98.93%	98.81%
a,b,c,d,→B	92.14%	93.92%	95.10%	97.62%	97.00%	96.74%
a,b,c,d,→C	82.2%	91.04%	86.4%	88.19%	99.20%	99.75%
a,b,c,d,→D	96.34%	98.26%	97.30%	97.43%	97.56%	97.46%
a,b,c,d,→E	93.65%	92.52%	96.34%	96.23%	96.97%	97.96%
平均值	91.43%	93.60%	94.18%	95.53%	97.93%	98.14%

为了更好地反映跨网实验的表现情况，将其与同网实验进行了对比，并设计了如下两类对比实验。

第一类对比实验：同网实验与跨网实验在不同数据集上的实验结果如图 4 所示。图 4（a）中同网 ISCX2012AS 表示仅使用同一个网络下的数



(a) 在ISCX2012AS上的实验结果



(b) 在CICIDS2017AS上的实验结果

图 4　同网实验与跨网实验在不同数据集上的实验结果

据进行训练和测试，即仅使用 ISCX2012AS 数据集作为训练和测试数据；跨网 CICIDS2017→ISCX2012AS 则表示使用 CICIDS2017AS 的数据进行训练，用于检测 ISCX2012AS 的数据，图 4（b）与之类似。通过改变 K 的取值来反映样本数量对准确率及检测率的影响。

第二类对比实验：同网实验与跨网实验结果的分布如图 5 所示。

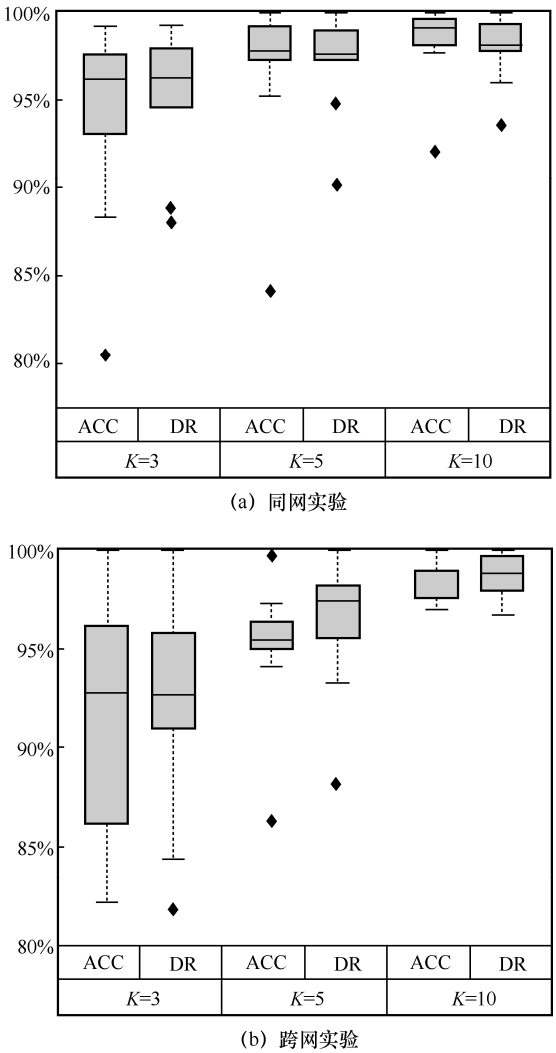


图 5　同网实验与跨网实验结果的分布

通过表 6 和表 7、图 4 和图 5，可以得出以下两个结论。

（1）本文方法具有较好的跨网适应能力。即使当 K=3 时，在两种数据集上的平均检测率仍能够达到 90.51%、93.60%，并且与同网实验一样，随着样本数量的增加，ACC 和 DR 也在逐步上升。当 K=10 时，在跨网实验下的检测结果与同网实验下的检测结果基本持平，最大波动幅度不超过 1.15%，这也说明跨不同数据集的训练和测试是可行的，从广义上来讲，它们都属于同一类型的计算机网络，其流量具有一定的共性。

（2）跨网实验在样本数量过少（如 K=3）时波动大，但随着样本数量的增加，跨网实验能有效减少异常点的出现。在图 5（a）中无论 K 的取值是多少，均有异常点，说明同网实验在预测某些攻击类型时较为欠缺，在预测这些攻击类型时准确率和检测率会大幅降低。而在图 5（b）中，异常点的数量明显少于图 5（a），当 K=10 时，异常点已经消失，说明跨网实验能有效减少异常点的出现，有效解决误报率过高的问题。

4　比较和讨论

4.1　与同类工作的对比

小样本网络入侵检测是一个比较新的研究领域。据笔者所知，已有的可供比较的相关工作还不多，可使用的数据集也比较少，因此本文构建了可用的数据集。由于小样本网络入侵检测不同于传统的入侵检测，本文基于真实的网络流量（ISCX2012 数据集和 CICIDS2017 数据集）构建数据集。具体来说，利用两个公开的网络流量数据源构建小样本检测数据集并对所提方法进行评估。本文构建的小样本数据集 ISCX2012AS 与原始 ISCX2012 数据集共享相同的原始网络流量（CICIDS2017AS 和 CICIDS2017 同理）。因此，本文对最近使用 ISCX2012 或 CICIDS2017 数据集的几项研究进行了概述，需要说明的是，基于元学习的连续小样本入侵检测方法[23]使用额外的数据集 NDSec-1 对 CICIDS2017 的攻击类型进行了

补充。此外，一种新型的多阶段层次入侵检测方法[22]研究的是零日样本的情况，零日样本指的是尚未被公开披露或广泛知晓的安全漏洞或攻击技术，它们在被发现和利用之前很少被研究人员和安全专家接触到，零日样本通常被认为是小样本，本文选择与之进行对比。

本文方法和相关研究工作中的检测结果和样本数量对比见表 8，本文方法在 ISCX2012AS 数据集上的检测率最高可达 99.29%，在 CICIDS2017AS 数据集上的准确率最高可达 97.93%，相比目前已有的 FC-Net（在 CICIDS2017FS 数据集上的检测率为 99.17%）和基于元学习的连续小样本入侵检测方法（准确率为 97.56%）分别提高了 0.12%和 0.37%，优于表 8 中的其他方法。

4.2　训练轮次和更新次数对实验的影响

为了进一步探究本文方法的性能，对轮次（epoch）和更新次数（update）这两个超参数做进一步分析。设计如下实验，设置更新次数为 0～5，把更新 0 次作为不使用本文方法的对比实验，并设置轮次为 0～20。

轮次和更新次数对检测结果的影响如图 6 所示，对于更新次数来说，当更新 0 次时，准确率保持在 50%附近，符合普通二分类的特性；当更新次数变成 1、2、3 时，准确率显著上升；当更新次数变成 4、5 时，准确率与更新 3 次基本保持一致，为了符合网络入侵检测的时效性，本文认为经历了 3 次更新后模型已经达到拟合状态，虽然继续更新能使模型的准确率小幅上升，但增幅甚小，因此在本次实验中选取更新 3 次。此外，当更新次数为 3 时，迭代次数在 0～2 个轮次时，准确率大幅上升，在 2～10 个轮次时处于轻微波动，经过 10 个轮次之后基本平稳。

表 8　本文方法和相关研究工作的检测结果和样本数量对比

方法	数据集	样本数量/个	准确率	检测率
Deep-CapsNet 与 ARCN(2021 年)[26]	ISCX2012FS	5	N/A	78.35%
Deep-CapsNet 与 ARCN(2021 年)[26]	CICIDS2017FS	5	N/A	81.65%
FC-Net(2020 年)[13]	ISCX2012FS	5	97.56%	98.78%
FC-Net(2020 年)[13]	CICIDS2017FS	5	94.33%	99.17%
FS-IDS(2022 年)[15]	CICIDS2017	5	97.51%	99.00%
基于 L2F 和模型无关的入侵检测方法(2023 年)[20]	CICIDS2017	10	94.66%	96.68%
基于元学习的连续小样本入侵检测方法(2022 年)[23]	CICIDS2017+NDSec-1	10	97.56%	N/A
一种新型的多阶段层次入侵检测方法(2023 年)[22]	CICIDS2017	47（零日样本）	96.00%	95.75%
本文方法	ISCX2012AS	3	94.55%	95.17%
		5	97.93%	97.79%
		10	98.83%	99.29%
本文方法	CICIDS2017AS	3	93.38%	94.86%
		5	95.45%	96.47%
		10	97.93%	98.14%

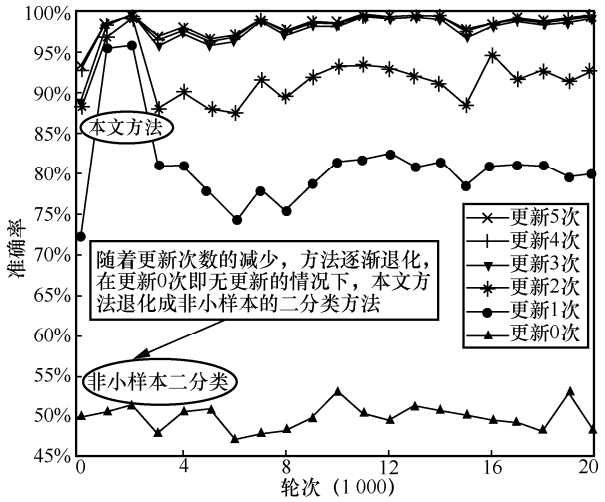


图 6　轮次和更新次数对检测结果的影响

4.3　误报率、漏报率指标分析

在实际应用时，小样本条件下的入侵检测除了考虑其实用性和准确率，还需要关注漏报率、误报率和样本数量对检测效果的影响。本文通过如下实验来分析上述指标。

首先，选择不同网络环境下的同网和跨网的实验结果，涵盖了不同网络环境和攻击类型的情况。其次，表 8 中的其他算法所提供的数据均来自单一网络环境下的检测结果，为了满足鲁棒性和泛化性，对不同网络环境下的实验结果求均值来模拟在真实的网络环境中的复杂情况，减少对特定网络环境的依赖，这样的评估方法更具有实际应用的价值，并能够更好地反映算法的整体性能。然后，将滑动窗口设置为 5 来计算窗口内数据的均值，以此降低噪声的影响，得到准确的变化趋势。

各指标随着样本数量的变化趋势如图7所示，样本数量为 5 时漏报率和误报率均低于 5%，样本数量为 10 时均低于 3%，样本数量为 15 时均低于 2%。样本数量为 1～5 时，准确率、检测率、精确率有明显的上升趋势；样本数量为 5～10 时，虽然上升趋势依然存在，但增长速度开始减慢；样本数量为 10～15 时，上升趋势较为缓慢。

通过对图 7 的分析，可以得到以下两个结论。

（1）本文提出的检测方法在样本数量为10时，漏报率和误报率均低于3%，其中漏报率为2.28%，已经可以达到实用要求。这说明即使面临样本数量有限的挑战，本文提出的检测方法仍然能够提供较高的准确性和可靠性。

（2）本文提出的检测方法在样本数量为 15 时的效果已达到最佳。样本数量为 15 时，误报率、漏报率均小于 2%，其中检测率为 98.33%。

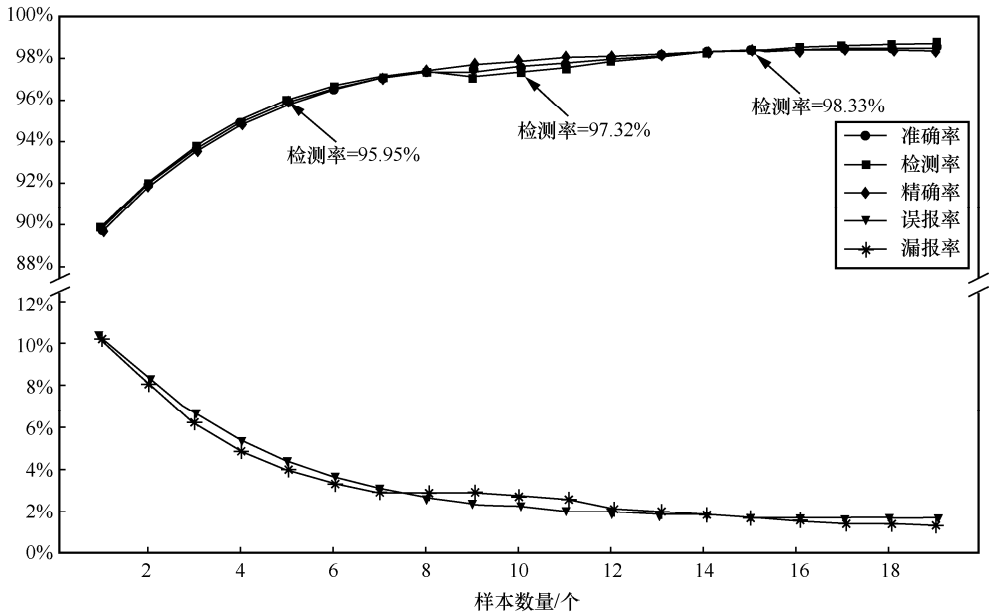


图 7　各指标随着样本数量的变化趋势

而且随着样本数量的增加，准确率、检测率、精确率对应的曲线上升缓慢，已经到达了饱和状态。该样本数量远小于非小样本条件下的一般网络入侵检测方法，进一步论证了本文工作的主要意义。

4.4　面对新型攻击类型

对于网络入侵检测领域而言，ISCX2012 和 CICIDS2017 数据集是经典且有代表性的数据集。它们被广泛应用于研究和评估入侵检测算法的性能。然而，由于网络环境的不断演变和新型攻击的不断出现，这些经典数据集在涵盖新型攻击方面存在一定的局限性。为了扩展实验研究的范围并更好地应对新型攻击，本文额外引入了 CICIDS2018 数据集，增加了新型的 DoS 攻击、DDoS 攻击、僵尸网络、暴力破解等。

根据第 4.3 节的实验结果，为满足高检测率、小样本的条件限制，选择 5～10 作为本次实验的样本数量范围。此外，控制单一变量，仅数据集不同，其余参数保持一致，包含新型攻击类型的检测结果如图 8 所示。

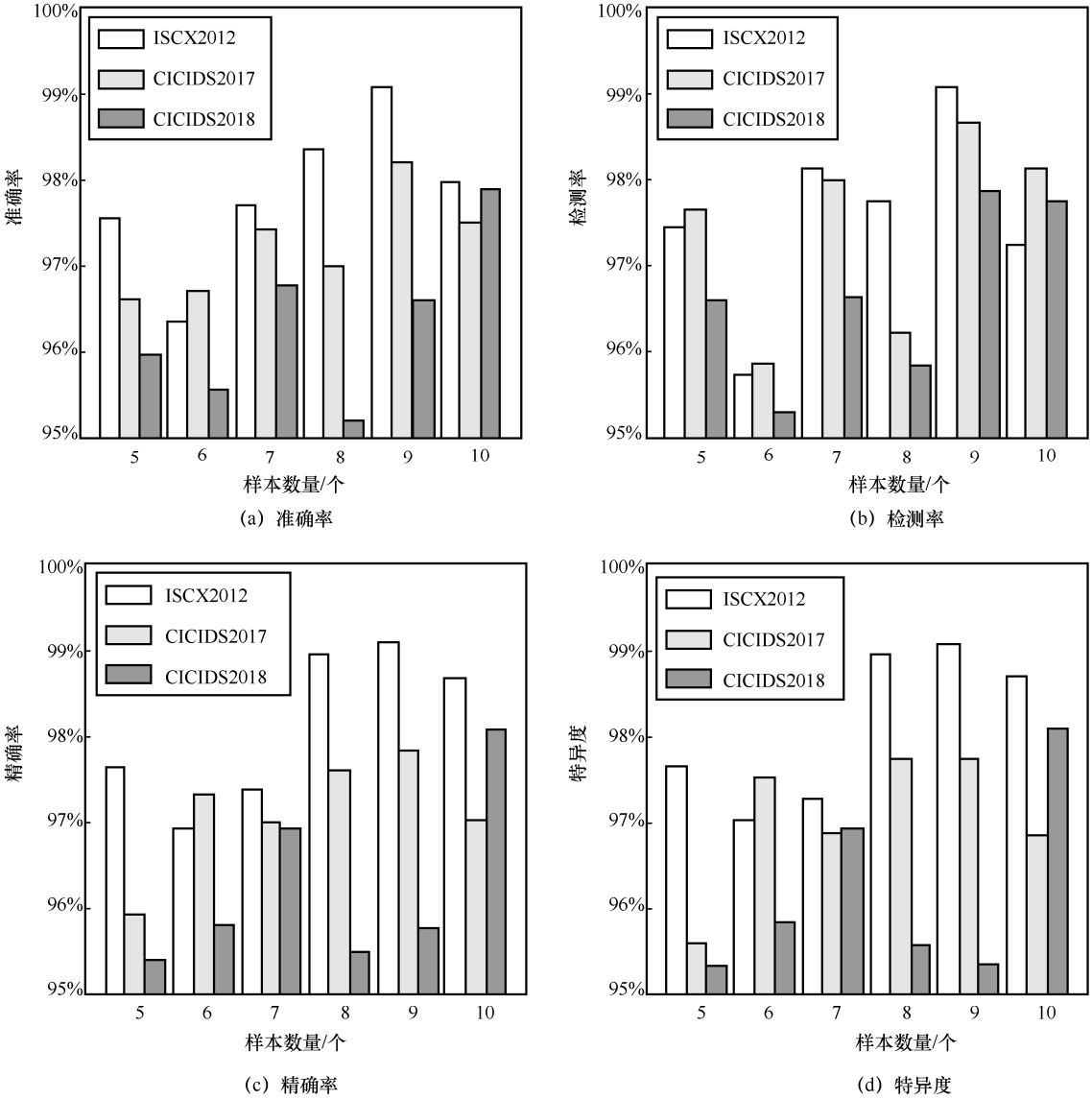


图 8　包含新型攻击类型的检测结果

通过图 8 可以发现，CICIDS2018 在样本数量为 10 时的各项指标均达到了 97%，与 ISCX2012 的检测结果差值不超过 1%。并且准确率、精确率、特异度均高于 CICIDS2017。这说明了本文方法在面临新型的网络环境及数据集的情况下依然有较高的可靠性和准确性，即使是在样本数量不充足的情况下，各指标也能达到 97%，因此本文方法不依赖于特定的网络环境和样本数量。

此外，新型攻击层出不穷，面对新型攻击类型，一种最优的评估方法应当引入最新采集的真实网络流量数据。限于研究条件，本文提出的检测方法还有待进一步在正式网络流量数据上进行验证。

4.5 局限性

首先，本文中的数据类型的种类过少，导致本文方法目前只适用于二分类问题，即只能检测出正常数据或入侵数据，而不能分析出具体是哪一种入侵。其次，本文在框架设计上，使用了基于元学习的算法，因此在任务的选取上必须具有一定的关联性，如果出现类型差别过大的入侵数据，则会导致算法的准确率大幅降低。为了解决这些问题，未来的研究可以考虑使用数据类别增强算法增加样本类型数量，以使本文方法能处理多类型任务。同时，考虑采用更优秀的框架设计，以有效处理关联性较弱的任务。这些改进将是未来研究的重点。

5 结束语

针对小样本条件下的网络入侵检测准确率低的问题，采用元学习的思想设计了多重循环的算法结构，提出了一种适用于小样本条件的网络入侵检测方法，大幅减少了训练时间，达到了较高的检测率。针对小样本条件下入侵检测数据集缺乏的问题，本文使用公开数据集的 pcap 包构建了 ISCX2012AS 与 CICIDS2017AS 两个数据集，该处理方法可使模型学习到更多有效特征。为了验证本文方法的有效性，本文做了大量的实验，并与多个同类工作进行比较。实验结果表明，面对 ISCX2012 与 CICIDS2017 数据集，本文方法在更加严格的小样本条件下仍然具有更优的性能，在 10 个训练样本下的检测率最高可达 99.29%，准确率最高可达 97.93%，相比目前已有算法分别提升了 0.12% 和 0.37%。此外，还引入 CICIDS2018 的数据集来更好地应对新型攻击，实验结果表明，在 10 个训练样本的条件下，本文方法的检测率已经超过了 97%，并且准确率、精确率、特异度指标均高于在 CICIDS2017 数据集上的检测结果。在后续研究中，将对本文方法进行优化，以提高其在单样本条件及任务关联性不强的环境中的准确率。

参考文献：

[1] LEE S W, SIDQI H M, MOHAMMADI M, et al. Towards secure intrusion detection systems using deep learning techniques: comprehensive analysis and review[J]. Journal of Network and Computer Applications, 2021(187): 103111.

[2] ZHANG Y, LI G Q, DUAN Q Q, et al. An interpretable intrusion detection method based on few-shot learning in cloud-ground interconnection[J]. Physical Communication, 2022(55): 101931.

[3] LI W H, LIU X L, BILEN H. Cross-domain few-shot learning with task-specific adapters[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 7151-7160.

[4] ZHANG Z Z, LAN C L, ZENG W J, et al. Uncertainty-aware few-shot image classification[C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2021: 3420-3426.

[5] AFRASIYABI A, LALONDE J F, GAGNÉ C. Mixture-based feature space learning for few-shot image classification[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2022: 9021-9031.

[6] KANG D, KWON H, MIN J H, et al. Relational embedding for

few-shot classification[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2022: 8802-8813.

[7] ALDWAIRI T, PERERA D, NOVOTNY M. An evaluation of the performance of restricted Boltzmann machines as a model for anomaly network intrusion detection[J]. Computer Networks, 2018(144): 111-119.

[8] ABDELMOUMIN G, RAWAT D B, RAHMAN A. On the performance of machine learning models for anomaly-based intelligent intrusion detection systems for the Internet of things[J]. IEEE Internet of Things Journal, 2022, 9(6): 4280-4290.

[9] HAGHIGHAT M H, LI J. Intrusion detection system using voting-based neural network[J]. Tsinghua Science and Technology, 2021, 26(4): 484-495.

[10] BASATI A, FAGHIH M M. DFE: efficient IoT network intrusion detection using deep feature extraction[J]. Neural Computing and Applications, 2022, 34(18): 15175-15195.

[11] SOLTANI M, SIAVOSHANI M J, JAHANGIR A H. A content-based deep intrusion detection system[J].International Journal of Information Security, 2022, 21(3): 547-562.

[12] LIANG W, HU Y Y, ZHOU X K, et al. Variational few-shot learning for microservice-oriented intrusion detection in distributed industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2021, 18(8): 5087-5095.

[13] XU C Y, SHEN J Z, DU X. A method of few-shot network intrusion detection based on meta-learning framework[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3540-3552.

[14] ILIYASU A S, ABDURRAHMAN U A, ZHENG L R. Few-shot network intrusion detection using discriminative representation learning with supervised autoencoder[J]. Applied Sciences, 2022, 12(5): 2351.

[15] YANG J C, LI H W, SHAO S, et al. FS-IDS: a framework for intrusion detection based on few-shot learning[J]. Computers & Security, 2022, 122: 102899.

[16] OUYANG Y K, LI B B, KONG Q L, et al. FS-IDS: a novel few-shot learning based intrusion detection system for SCADA networks[C]//Proceedings of ICC 2021 - IEEE International Conference on Communications. Piscataway: IEEE Press, 2021: 1-6.

[17] YU L, DONG J T, CHEN L H, et al. PBCNN: packet bytes-based convolutional neural network for network intrusion detection[J]. Computer Networks, 2021(194): 108117.

[18] WANG Z M, TIAN J Y, QIN J, et al. A few-shot learning-based Siamese capsule network for intrusion detection with imbalanced training data[J]. Computational Intelligence and Neuroscience, 2021: 1-17.

[19] GAMAL M, ABBAS H M, MOUSTAFA N, et al. Few-shot learning for discovering anomalous behaviors in edge networks[J]. Computers, Materials & Continua, 2021, 69(2): 1823-1837.

[20] SHI Z X, XING M Y, ZHANG J, et al. Few-shot network intrusion detection based on model-agnostic meta-learning with L2F method[C]//Proceedings of 2023 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2023: 1-6.

[21] YE T P, LI G L, AHMAD I, et al. FLAG: few-shot latent Dirichlet generative learning for semantic-aware traffic detection[J]. IEEE Transactions on Network and Service Management, 2022, 19(1): 73-88.

[22] VERKERKEN M, D'HOOGE L, SUDYANA D, et al. A novel multi-stage approach for hierarchical intrusion detection[J]. IEEE Transactions on Network and Service Management, 2023, PP(99): 1.

[23] XU H, WANG Y J. A continual few-shot learning method via meta-learning for intrusion detection[C]//Proceedings of 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT). Piscataway: IEEE Press, 2022: 1188-1194.

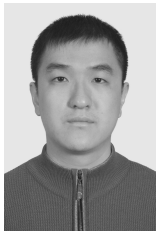
[24] SHARAFALDIN I, HABIBI LASHKARI A, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]//Proceedings of the 4th International Conference on Information Systems Security and Privacy. San Francisco: Science and Technology Publications, 2018: 108-116.

[25] SHIRAVI A, SHIRAVI H, TAVALLAEE M, et al. Toward developing a systematic approach to generate benchmark datasets for intrusion detection[J]. Computers & Security, 2012, 31(3): 357-374.

[26] MA W G, ZHANG Y D, GUO J, et al. Few-shot abnormal network traffic detection based on multi-scale deep-CapsNet and adversarial reconstruction[J].International Journal of Computational Intelligence Systems, 2021, 14(1): 1-25.

[作者简介]

胡炜晨（2000− ），男，嘉兴学院信息科学与工程学院在读，主要研究方向为网络安全和机器学习。



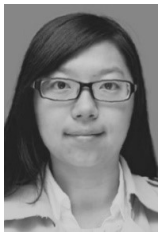
许聪源（1990− ），男，博士，嘉兴学院信息科学与工程学院讲师，主要研究方向为网络空间安全和智能信息处理。

刘思情（2002− ），男，嘉兴学院信息科学与工程学院在读，主要研究方向为人工智能和漏洞检测。

詹勇（2002− ），男，嘉兴学院信息科学与工程学院在读，主要研究方向为信息安全和深度学习。

王志强（2003− ），男，嘉兴学院信息科学与工程学院在读，主要研究方向为网络安全与人工智能。

陈广辉（2002− ），男，嘉兴学院信息科学与工程学院在读，主要研究方向为人工智能和信息安全。



王晓琳（1989− ），女，博士，嘉兴学院信息科学与工程学院讲师，主要研究方向为智能回归测试和深度学习。