面向网络入侵检测的深度模型对比与展望

叶健，刘坤
湖北交通职业技术学院，湖北 武汉 430202
DOI: 10.61369/RTED.2026080008

摘　　要： 随着网络攻击越来越智能化、隐蔽化、规模化，传统的以特征匹配为基础的入侵检测技术已经不能有效地对抗未知的攻击、变种的攻击和高级持续性威胁。深度学习依靠自动特征提取、非线性建模以及高维数据处理，成为网络入侵检测技术的主要发展方向。本文对卷积神经网络、长短期记忆网络、Transformer、生成对抗网络以及混合深度学习模型的原理和入侵检测的适配性进行梳理，分析各个模型的技术优势、固有的缺陷和适用范围。在此基础上，指出目前深度学习入侵检测存在数据不平衡、可解释性差等主要问题，并提出轻量化、可解释、联邦化的未来发展方向，给入侵检测系统的建立和改善赋予实践参照。

关 键 词： 深度学习；网络入侵检测；卷积神经网络；Transformer；模型对比；网络安全

Comparative Analysis of Network Intrusion Detection Models Based on Deep Learning

Ye Jian, Liu Kun
Hubei Communications Technical College, Wuhan, Hubei 430202

Abstract: As network attacks become increasingly intelligent, covert and large-scale, traditional signature-matching based intrusion detection technologies can no longer effectively defend against unknown attacks, variant attacks and advanced persistent threats. With the capabilities of automatic feature extraction, nonlinear modeling and high-dimensional data processing, deep learning has become a major development trend of network intrusion detection technologies. This paper expounds the principles and adaptability of Convolutional Neural Network, Long Short-Term Memory, Transformer, Generative Adversarial Network and hybrid deep learning models for intrusion detection, and analyzes their technical advantages, inherent limitations and application scopes. Furthermore, it points out major problems existing in current deep learning-based intrusion detection, such as data imbalance and poor interpretability, and proposes future development directions including lightweight design, interpretability and federated learning. The research provides practical references for the construction and optimization of intrusion detection systems.

Keywords: deep learning; network intrusion detection; convolutional neural network; transformer; model comparison; network security

一、网络入侵检测的重要性

伴随着互联网技术的迅速普及以及数字业务的不断深入发展，网络空间已经成为了各个行业运转的基础性基础设施，大量的敏感业务数据以及用户信息都在网络中进行传输和存储，网络安全直接牵涉到个人权益、企业发展甚至国家安全。网络入侵行为属于威胁网络安全的重要因素，近几年来，勒索攻击、数据泄露、高级持续性威胁等攻击事件屡有发生，攻击方式不断更新换代，不但会造成业务停机，还会造成难以挽回的损失和社会影响[1]。

网络入侵检测属于网络安全防御体系的重要部分，它可以通过对网络传输数据的即时监测来察觉到异常访问以及恶意攻击行为，从而给安全处置赋予有力的支撑，这是创建主动防御体系的重要保障。传统的入侵检测技术依靠人工标注的攻击特征库来进行匹配检测，对于不断出现的零日漏洞、变种攻击以及隐蔽性强的高级持续性威胁来说，不但特征更新迟缓，而且很难对海量高维的网络流量数据展开处理，检测漏报率和误报率一直居高不下，已经不能适应目前复杂网络环境下的安全防御要求了。因此，探寻契合现代网络环境的高性能入侵检测技术，对于加强整体网络安全防御水平，保证各类数字化业务正常、安全地运转有着十分重大的现实意义[2]。

二、常见深度学习入侵检测模型对比分析

（一）卷积神经网络（CNN）

卷积神经网络的核心就是卷积层和池化层，用局部感受野和权值共享来提取数据的空间特征。一维 CNN 把一维流量特征向量直接当作输入，用一维卷积核滑动来提取局部特征，没有因为格式转换而造成信息的丢失，二维 CNN 把原始流量字节变成灰度图像，依靠已有的图像识别技术来检测攻击，特别适合于加密流量

的检测[3]。

由于传统的 CNN 对于时序特征的提取能力较差，因此本文使用时间卷积网络 (TCN) 来提高并行计算的同时保留时序依赖[4]。CNN 的核心优势就是计算效率高、参数少、训练速度快，单条流量推理时间可以低至0.02ms，对硬件资源要求低。它的不足之处在于不能很好地识别出超长时间依赖的行为，对于多个阶段的 APT 攻击也难以检测出来。CNN 在资源匮乏的边缘设备、物联网网关上被广泛地用作实时检测已知攻击和简单的异常流量。

（二）长短期记忆网络（LSTM）

长短期记忆网络是循环神经网络的一种改进，它使用了输入门、遗忘门和输出门这三个门控机制来克服传统 RNN 的梯度消失问题，并且可以很好地处理长序列时间依赖性[5]。入侵检测当中，LSTM 把网络流量当作时间序列，逐帧输入到网络里，依靠门控来挑选历史攻击信息，从而掌握流量的时序改变趋向。

双向 LSTM（BiLSTM）会同时对待测流量序列进行正向和反向处理，从而可以获取到攻击行为前后关联的信息，对于检测多阶段攻击来说比较合适，而 GRU 则是 LSTM 的一种简化形式，它把输入门和遗忘门合二为一，使得训练速度提高了大约30%，但是检测精度有所降低[6]。LSTM 的主要优势就是对于时序特征的准确把握，对于 DDoS、端口扫描这些连续型攻击的检测准确率可以达到92% 以上。它的缺点是并行计算能力弱，训练速度慢，不能处理超长的流量序列。LSTM 主要用在核心网络的入侵检测系统中，对具有时序特征的攻击行为进行检测。

（三）Transformer 模型

transformer 用自注意力机制来并行处理序列数据，并且可以捕捉到全局长距离依赖，完全解决了 rnn 系列模型的并行计算问题[7]。入侵检测的时候，Transformer 先用位置编码给流量序列加上时序信息，然后依靠自注意力机制算出每一个特征同其他所有的特征之间的联系权重，从而自动聚焦于和攻击有关的重要特征，忽略掉无用的噪音。

针对 Transformer 计算量大的问题，DistilBERT 通过知识蒸馏压缩模型参数，MobileViT 结合 CNN 与 Transformer 的优势，在保持精度的同时大幅降低计算量。Transformer 的核心优势就是可以捕捉到超长距离的时序依赖，对于周期长达数周的 APT 攻击检测效果较好，且具有较强的并行计算能力，训练速度比 LSTM 快3到5倍。其不足之处在于参数量大，基础模型参数达到数千万级，对于 GPU 资源的需求较高，不能直接部署到边缘设备上。Transformer 主要部署在数据中心、核心网络服务器上，用来检测复杂的未知攻击以及 APT 攻击。

（四）生成对抗网络（GAN）

生成对抗网络是由生成器和判别器组成的，用对抗博弈的方式来训练模型[8]。生成器学习正常流量概率分布，生成逼真的假样本；判别器分辨真实样本和生成样本，两者互相对抗、共同提高。GAN 主要完成两个任务，一是处理数据不平衡问题，用生成器创造出少量的攻击样本来增加训练集，提高对于罕见攻击的检测准确率，二是做无监督的异常检测，用判别器找出和正常流有不同之处的流量，无需大量的标注数据。

传统的 GAN 存在着训练不稳定、模式崩溃等缺点，WGAN 用 Wasserstein 距离解决了梯度消失的问题，WGAN-GP 又用梯度惩罚来提高训练的稳定性[9]。GAN 的主要优点就是可以有效地解决数据不平衡的问题，并且可以进行无监督学习，在标注数据不足的情况下也可以使用。它的缺点在于训练过程复杂，超参数的调节比较困难，生成出来的样本的真实性、多样性还需要提高。GAN 主要用在零日攻击检测和工业互联网场景中。

（五）混合深度学习模型

单一深度学习模型不能同时提取空间和时序特征，所以混合模型成了目前的研究热点[10]。混合模型将不同模型的优点结合起来以达到性能互补的目的，常见的有三种，分别是 CNN-LSTM 模型，先用 CNN 提取局部空间特征，再输入 LSTM 提取时序特征，兼顾高效性和时序捕捉能力；CNN-Transformer 模型，用 CNN 提取局部特征，用 Transformer 捕捉全局依赖，适合检测复杂的多阶段攻击；注意力增强混合模型，在基础模型的基础上加入注意力机制，自动关注重要的攻击特征，从而提高检测精度。

混合模型的核心优势就是综合性能最优，在主流数据集上 F1 值能达到93% 以上。其缺点就是模型结构复杂、计算量大、训练和部署难度高。混合模型适合于对检测精度要求较高的关键信息基础设施防护场景。

三、现存挑战与未来发展方向

（一）现存挑战

1. 数据不平衡和标注不足的问题，真实网络中正常流量和攻击流量的比例严重失衡，很少见的攻击样本只占很小的比例，造成模型检测少数类攻击的效果非常差。网络攻击数据标注成本高、周期长，而且会涉及到用户的隐私以及企业的机密，所以很难得到大量的高质量标注数据集。虽然 GAN 可以生成攻击样本，但是生成的样本真实性、多样性还不能满足实际需要，不能完全代替真实的标注数据。

2. 模型不可解释性强，深度学习模型属于典型的黑箱模型，不能解释决策的过程和依据，不能满足安全运维人员溯源取证的要求，也不能符合等保2.0等安全合规的要求。当模型出现误报或者漏报的时候，不能迅速找到问题的原因，不利于模型的优化和调试，也降低了安全人员对模型的信任度。

3、泛化能力以及域适应能力欠缺，单一数据集训练出来的模型，在别的网络环境里会表现出非常明显的性能下降情况，这是由于各个网络流量分布存在较大差别所造成的。模型很容易出现过拟合现象，在训练数据局部特征上拟合得很准，但是不能适应网络环境的变化以及攻击方式的更新，造成实际部署时检测效果不如实验环境。

4. 计算资源消耗和部署问题，Transformer 以及大型混合模型的参数量很大，训练和推理需要大量的 GPU 资源，在资源受限的边缘设备、物联网终端和嵌入式系统上无法部署使用，从而影响了其在分布式入侵检测中的应用。模型推理延迟过大，不能满足高并发网络实时检测的需求。

5. 对抗样本攻击威胁，攻击者只需要改变少量的流量特征就可以使深度学习模型误判为正常流量，对抗样本攻击具有隐蔽性好、成功率高这两个特点。目前对于深度学习入侵检测的对抗防御技术还不是很成熟，很难抵挡住复杂的对抗攻击，严重危害到入侵检测系统的安全。

（二）未来发展方向

轻量化深度学习模型就是利用模型剪枝、量化、知识蒸馏、神经架构搜索等技术，减少模型参数和计算量，使得模型可以部署到边缘设备和物联网终端上，从而实现分布式入侵检测。重点发展轻量级的 Transformer、轻量化的混合模型，在保证检测精度的基础上大幅度减少资源消耗。

2. 可解释人工智能（XAI），即把可解释技术与入侵检测模型结合起来，使用 LIME、SHAP 等方法来解释特征的重要性以及决策的过程，从而创建可视化的模型解释工具，提高模型的可信度和可维护性。另外还要建立可解释性的评价标准，促进可解释的入侵检测模型工程化应用。

3、联邦学习和隐私计算，在保证数据隐私的前提下，多个机构一起训练全局入侵检测模型来解决数据孤岛的问题，提高模型的泛化能力。采用差分隐私、同态加密等手段保证数据在训练和推理的过程中是安全的、隐私的。

4）多模态融合检测，将网络流量、系统日志、用户行为、威胁情报等各种不同的数据源进行融合，从不同的角度来刻画攻击行为，提高对复杂的攻击的检测能力。研究多模态数据对齐、融合、特征交互技术来解决多源数据异构性问题。

5. 自适应在线学习，即可以在线学习新的攻击特征的自适应模型，采用增量学习、终身学习等技术不断更新模型参数，适应不断变化的网络环境和攻击手段，使模型不断进化，不会因为频繁的离线重新训练而影响到系统的运行。

四、结语

深度学习技术给网络入侵检测带来革命性的变革，不同的模型在技术特点、性能表现和适用场景上各有所长。CNN 由于计算能力强，被用作边缘实时检测的首选，LSTM 对于时序特征明显、连续的攻击比较适合，Transformer 对于复杂的长周期 APT 攻击检测效果好，GAN 可以很好地解决数据不平衡的问题，混合模型利用优势互补来达到最佳的综合性能。虽然目前深度学习入侵检测仍然存在数据不平衡、可解释性差、泛化能力弱等诸多问题，但是随着轻量化、可解释、联邦化等技术的发展，它必将会在网络安全防护中发挥出越来越重要的作用。未来深度学习入侵检测会向着更加高效、更加可信、更加智能、更加协同的方向发展，给创建安全可靠的网络空间赋予有力支持，保证数字经济的健康发展。

参考文献

[1] 宋亚龙 . 基于深度学习的物联网网络入侵检测方法研究 [D]. 郑州大学 ,2024.

[2] 周文刚 . 基于双向生成对抗网络的网络入侵检测方法 [D]. 长安大学 ,2024.

[3] 朱鹏 . 基于特征选择与深度学习的网络入侵检测模型研究 [D]. 辽宁科技大学 ,2024.

[4] 杨晓文 , 张健 , 况立群 , 等 . 融合 CNN-BiGRU 和注意力机制的网络入侵检测模型 [J]. 信息安全研究 ,2024,10(3):202-208.

[5] 杨超 , 杜琪琪 , 范波 , 等 . 深度学习下时序特征提取的网络入侵检测模型 [J]. 保密科学技术 ,2024,(1):51-60.

[6] 吴瑕 . 深度学习在网络入侵检测系统中的实用性与效率研究 [J]. 信息记录材料 ,2024,25(1):222-224.

[7] 王瀚文 . 基于自编码器和 GoogLeNet 的入侵检测研究 [D]. 辽宁工程技术大学 ,2023.

[8] 杜佳玮 . 基于联邦学习的网络入侵检测技术研究 [D]. 西京学院 ,2023.

[9] 谷文杰 . 基于深度学习的网络入侵检测方法研究 [D]. 西京学院 ,2023.

[10] 白文荣 , 马琳娟 , 巩政 . 基于菌群优化深度学习的网络入侵检测模型 [J]. 南京理工大学学报 ,2023,47(5):636-642.

