文章类型：论文|刊号（ISSN）：2737-4505(P) / 2737-4513(O)

基于深度学习的网络入侵检测系统问题研究

吴赵本卓　孙世杰　卞延江　杜程前
安徽信息工程学院
DOI:10.32629/etd.v6i5.16874

[摘　要] 本文旨在研究基于深度学习的网络入侵检测系统中的关键问题。通过系统分析深度学习在入侵检测中的应用现状,指出当前系统在检测准确性、实时性和泛化能力三方面存在的突出问题,并提出针对性的优化对策。研究结果表明,通过集成注意力机制、模型轻量化、对抗训练等综合策略,能够有效提升系统性能。结论表明,未来需从模型创新、数据集构建和工程优化等多维度协同推进,才能实现高效可靠的智能入侵检测系统。
[关键词] 深度学习；网络入侵检测；系统优化；对抗训练
中图分类号：N945.15　文献标识码：A

Research on Key Issues in the Design of Network Intrusion Detection System Based on Deep Learning

Zhaobenzhuo Wu　Shijie Sun　Yanjiang Bian　Chengqian Du
Anhui University of Information Technology, Wuhu City, Anhui Province

[Abstract] This paper aims to study the key issues in the design of network intrusion detection systems based on deep learning. By systematically analyzing the current application status of deep learning in intrusion detection, it points out the prominent problems existing in the current system in terms of detection accuracy, real-time performance, and generalization ability, and proposes targeted optimization strategies. The research results show that by integrating attention mechanisms, model lightweighting, and adversarial training, the system performance can be effectively improved. The conclusion indicates that in the future, it is necessary to promote the development of efficient and reliable intelligent intrusion detection systems from multiple dimensions such as model innovation, dataset construction, and engineering optimization.
[Key words] Deep Learning; Network Intrusion Detection; System Optimization; Adversarial Training

前言

近年来,随着数字化转型加速和网络威胁的日益复杂化,网络安全已上升为国家战略层面的核心议题。为应对高级持续性威胁、零日攻击等新型风险,我国相继出台《网络安全法》、《关键信息基础设施安全保护条例》等重磅文件,明确提出要强化态势感知与入侵监测能力。这些政策导向不仅凸显了提升网络防御体系的紧迫性,更对入侵检测技术的智能化革新提出了明确要求。在此背景下,传统基于签名的检测手段已显乏力,而深度学习技术因其在复杂模式识别上的巨大潜力,成为实现主动、智能安全防御的关键突破口。本研究旨在深入探讨基于深度学习的网络入侵检测系统,直面其在准确性、实时性等方面的挑战,以响应国家政策号召,为构建自主可控的先进网络安全防护体系提供理论支撑与实践路径。

1 深度学习在网络入侵检测中的应用现状

1.1基于自编码器的异常检测模型

基于自编码器的异常检测模型作为一种典型的无监督学习方法,其核心优势在于无需依赖大量已标注的攻击样本,仅通过大量正常网络流量数据即可完成自我训练,从而能够有效应对层出不穷的未知攻击类型。该模型通常由编码器和解码器两部分构成,编码器负责将高维的原始输入数据(如网络流统计特征、数据包字节序列等)压缩到一个低维的潜空间表示中,以捕捉其最本质的特征;随后,解码器则致力于从这个压缩的潜表示中尽可能地重构出原始输入数据。在模型训练阶段,通过最小化正常流量重构误差的损失函数,使得自编码器能够精准学习正常网络行为的稳定模式与分布特征。当进入实际检测阶段,输入待分析的网络流量时,若该流量行为模式与学习到的正常模式高度一致,则模型能够以较低的重构误差准确还原；反之,若输入的是异常或攻击流量,由于其数据分布与训练所用的正常数据存在显著差异,模型将难以进行精准重构,从而导致重构误差急剧升高。最终,系统通过设定一个合理的误差阈值,一旦某次

流量的重构误差超出该阈值,即可将其判定为异常入侵行为,实现了对未知威胁的有效探测,展现了其在零日攻击检测方面的独特潜力。

1.2基于循环神经网络的时序流量分析

网络入侵行为往往并非孤立事件,而是在时间维度上展现出连续的、具有前后关联性的序列模式,基于循环神经网络的时序流量分析方法正是为了捕捉这种动态依赖关系而设计的。与传统的前馈神经网络不同,RNN通过其内部隐藏状态的循环连接,具备了一种“记忆”能力,能够将之前时间步的信息传递到当前的计算中,从而非常适合于处理网络连接序列、系统调用链或持续的网络会话流等时序数据。特别是长短期记忆网络和门控循环单元这两种先进的RNN变体,通过引入精巧的门控机制,有效解决了传统RNN在训练过程中可能出现的梯度消失或爆炸问题,使其能够学习并记忆更长距离的时间依赖关系。在实际应用中,模型通过分析连续的网络事件,例如,它可以学习到一个合法的用户登录会话中“TCP三次握手”、“SSL/TLS协商”、“HTTP请求”这一系列事件所应遵循的正常时序与逻辑关系。当出现异常行为,如分布式拒绝服务攻击所引发的在极短时间内来自大量不同源IP的TCP SYN连接请求洪泛,或是暴力破解攻击导致的连续且高频的失败认证尝试,这些行为序列模式会显著偏离模型所学到的正常时序规律,从而被RNN模型精准地识别为入侵信号,极大地提升了对慢速扫描、高级持续性威胁等具有强时序特征的复杂攻击的检测能力。

1.3基于卷积神经网络的空间特征提取

基于卷积神经网络的空间特征提取方法将网络数据视为特殊形式的″图像″,通过其强大的空间特征学习能力实现入侵检测。该方法通过将网络流量数据重构为二维矩阵形式,如将数据包载荷按字节排列或将会话特征组合成特征图,进而应用CNN进行深度分析。CNN通过卷积层的局部连接和权值共享特性,使用多个可学习的滤波器在输入数据上滑动,自动提取从低级到高级的层次化特征。随后的池化层则通过下采样操作保留显著特征,增强模型的平移不变性。这种端到端的特征学习方式使CNN能够有效识别恶意软件通信的固定模式、特定攻击的载荷特征等具有空间局部相关性的攻击签名,在已知攻击变种检测和基于内容分析的入侵识别方面表现出色。

2 网络入侵检测系统存在的问题

2.1检测准确性不足

检测准确性不足是基于深度学习的网络入侵检测系统面临的核心瓶颈,其根源在于多重因素的复杂交织。首要挑战是网络环境中固有的、极端的数据不平衡问题,正常流量在数量上往往占据绝对主导,而真正的攻击样本,尤其是新型或复杂的威胁,在训练数据中占比极低,这导致模型在优化过程中会自然地倾向于将多数类(即正常流量)分类正确,而对少数类(即攻击流量)的识别能力则严重不足,造成高漏报率,使得攻击得以“隐身”于海量正常背景流量中。其次,尽管深度学习能够自动提取特征,但其性能依然高度依赖于输入数据的质量,原始网络流量中充斥着大量无关噪声、冗余特征以及由于数据采集或预处理不当引入的偏差,这些“脏数据”会严重误导模型的学习过程,使其难以捕捉到攻击行为的本质判别性特征。此外,深度学习模型本身如同一个“黑箱”,其网络结构深度、神经元数量、激活函数选择以及学习率等超参数的组合极为复杂,缺乏系统性的理论指导来直接确定最优配置,不恰当的参数设置极易导致模型陷入对训练数据的过拟合,即完美“记住”了训练集甚至包括其中的噪声,却丧失了泛化能力,或者在相反情况下出现欠拟合,无法从数据中学到有效模式,这两种情况都会在真实世界的复杂流量面前导致灾难性的准确率下降。

2.2系统实时性差

系统实时性差是阻碍基于深度学习的网络入侵检测系统在实际高带宽网络环境中部署应用的关键障碍,其瓶颈主要体现在模型计算与数据处理两个层面。在模型计算层面,诸如深度卷积神经网络或复杂循环神经网络在内的先进深度学习模型,通常拥有数百万乃至数十亿的参数量,进行一次前向推理需要执行海量的浮点矩阵运算,这对于需要实现线速处理、微秒级延迟响应的骨干网或数据中心入口而言,构成了巨大的计算压力,极易在高吞吐量流量冲击下造成数据包积压、处理延迟甚至丢包,使得入侵检测系统本身成为网络性能的瓶颈。在数据处理层面,从原始数据包到模型可接受的规整化张量输入,需要经历一个复杂且耗时的预处理流水线,包括数据包捕获、协议解析、流特征聚合、特征值标准化或归一化等一系列步骤,这些操作同样需要消耗可观的CPU计算资源和内存带宽,其总耗时在某些场景下甚至可能超过模型推理本身,进一步加剧了系统整体延迟。此外,现有的许多研究模型为了追求在基准数据集上的高精度,往往在设计上趋于复杂和庞大,缺乏对实际部署环境中苛刻的计算与时间约束的考量,从而导致了“实验室精度高,工程落地难”的窘境,无法满足现代网络安全对威胁实时响应和即时阻断的刚性需求。

2.3模型的泛化能力差

模型的泛化能力差是基于深度学习的网络入侵检测系统从理论研究迈向实际部署过程中所暴露出的最严峻挑战之一,其本质在于模型在特定环境下学到的“知识”难以适应动态变化的真实网络世界。泛化能力差的首要原因是广泛存在的数据集偏差,当前该领域的大量研究严重依赖于KDD CUP 99、NSL-KDD等历史悠久的基准数据集,这些数据集不仅年代久远,其包含的攻击类型和流量特征与当前真实网络中活跃的零日攻击、高级持续性威胁等复杂攻击模式已相去甚远,导致在其上精心调优的模型一旦部署到现实网络环境中,便会因数据分布的巨大差异而性能锐减。其次,深度学习模型自身在面对对抗性攻击时表现出令人担忧的脆弱性,攻击者可以利用模型的梯度信息,通过向恶意流量中注入人眼或传统检测方法难以察觉的微小扰动,即可成功地“欺骗”模型,使其将攻击流量错误地判定为正常,这种安全漏洞使得依赖深度学习模型的NIDS在面对有针对性的智能攻击时变得极不可靠。最后,网络环境本身就是一个持续演

化的动态系统,新的应用协议、用户行为模式以及网络配置变更都会导致正常流量的统计特征分布随时间发生漂移,一个在历史数据上训练出的静态模型无法适应这种变化,会逐渐将新的正常行为误判为异常,导致误报率攀升,或因无法识别新型攻击模式而导致漏报,最终使得系统失效。

3 网络入侵检测系统的对策

3.1提高检测准确性

提升检测准确性需从数据、模型和算法三个层面进行系统性优化。针对数据不平衡这一核心挑战,可采用合成少数类过采样技术等高级重采样方法,在特征空间内智能地生成具有代表性的少数类样本,从而平衡类别分布；同时,在算法层面引入焦点损失或加权交叉熵等代价敏感学习机制,通过在损失函数中为难以分类的少数类样本分配更高权重,迫使模型在训练过程中更加关注攻击流量。在模型架构方面,引入注意力机制是提升判别能力的有效途径,它能使模型像人类一样专注于输入序列或特征图中与攻击最相关的关键部分,例如在长流量序列中锁定异常连接请求,或在数据包负载中聚焦于恶意代码片段,从而有效抑制无关噪声的干扰。此外,采用模型集成策略,如将擅长捕捉空间特征的CNN、擅长时序分析的RNN以及对异常敏感的自编码器进行有机组合,通过投票或堆叠的方式汇聚“集体智慧”,能够综合利用不同模型的优势,显著提升整体检测的稳健性和精确度,有效应对多样化的攻击模式。

3.2提高系统实时性

为满足高速网络环境下的实时检测需求,必须着力于模型轻量化、硬件加速和流程优化。模型轻量化是核心,可通过一系列技术手段实现：采用网络剪枝移除模型中冗余的权重和神经元,大幅减少参数量和计算量；应用知识蒸馏技术,利用一个大型、精确的教师模型来指导一个小型学生模型的训练,使小模型在保持接近大模型性能的同时实现推理速度的飞跃；或直接选用专为高效计算设计的轻量级网络架构(如MobileNet、SqueezeNet)作为基础。在硬件层面,充分利用GPU、TPU乃至FPGA等专用处理器的并行计算能力,对模型推理过程进行极致加速,并将预处理与模型推理环节部署成高效的并行流水线,以隐藏处理延迟。同时,在流量处理策略上,可采用基于时间窗口的小批量流处理代替严格的单包处理,在保证检测时效性的前提下,通过微小延迟换取更高的整体吞吐量,从而确保系统在高负载下仍能稳定运行。

3.3改善模型的泛化能力

增强模型的泛化能力是确保入侵检测系统在真实复杂环境中保持效能的根本。首要任务是推动数据集的现代化与多元化,积极采用CIC-IDS2017、UNSW-NB15等包含当代攻击模式的新型公开数据集进行模型训练与评估,并鼓励在工业界构建更具代表性的真实网络流量基准,以从根本上减少训练数据与实战环境之间的分布偏差。为提升模型在面对恶意攻击时的鲁棒性,必须引入对抗性训练,在模型训练过程中主动地将其与精心生成的对抗样本进行对抗,使模型在学习区分正常与攻击模式的同时,也学会抵抗这些细微的、旨在欺骗模型的扰动,从而构建起更坚固的防御壁垒。最后,为了应对网络环境的动态演化,应部署在线学习或增量学习机制,使模型能够在不遗忘已有知识的前提下,持续地从新到达的网络流量中学习并适应新的正常行为模式和新兴威胁,实现模型的自我进化与终身学习,从而长期保持其在实际部署中的有效性和适应性。

4 结论

综上所述,深度学习技术为构建下一代智能入侵检测系统提供了强大支持,但其在检测准确性、实时性与泛化能力方面仍面临显著挑战。未来研究应着力于模型架构创新与轻量化设计,推动高质量数据集的构建与共享,并积极探索在线学习与对抗训练等自适应机制。通过算法优化、工程实现与安全保障的深度融合,才能推动基于深度学习的网络入侵检测系统突破现有瓶颈,最终构建出能够有效应对动态威胁的主动防御体系。

[参考文献]

[1]吴祖康.基于深度学习的网络入侵检测系统设计[J].网络安全技术与应用,2025,(10):64-66.

[2]叶勇飞,匡石磊,王冠.基于深度学习的网络入侵检测系统设计[J].软件,2025,46(03):54-56.

[3]朱悦云.基于深度学习的工业控制网络入侵检测系统设计[J].电子技术,2025,54(01):118-120.

作者简介：

吴赵本卓(2005--),男,汉族,安徽省芜湖市人,本科,研究方向：计算机科学与技术。