基于人工智能的网络入侵检测系统优化研究

杜秀玲

贝子府镇人民政府，内蒙古自治区赤峰市，024000；

摘要：面对越来越复杂隐蔽的网络攻击，传统的入侵检测技术由于其静态的规则库和有限的特征识别能力，很难应对新的攻击。本研究主要目的在于探究并构建起一种依靠人工智能，尤其是深度学习技术来改善网络入侵检测系统性能的方案。设计层次化的特征提取框架，使用改进的深度神经网络分类器，从大量的网络流量数据中自动、准确地识别出异常行为和恶意攻击。系统优化策略包含特征选择、模型结构、实时处理机制等各个方面。实验表明，所提出的方法在主流数据集上具有较好的检测精度和较低的误报率，可以有效提高网络安全防护系统的智能性、自适应性和实时响应能力，为下一代主动防御体系的建立提供一条可行的技术路径。

关键词：人工智能；入侵检测；深度学习；系统优化

DOI：10.69979/3060-8767.26.01.005

数字时代到来以后，网络空间成为国家安全和社会经济运转的基础载体，其所遭遇的安全威胁也越来越多样、复杂且持久。传统的基于固定签名、预先设定规则的入侵检测系统，对于零日攻击、高级持续性威胁、复杂的多变攻击行为等，其性能瓶颈就越来越突出。以深度学习为代表的人工智能技术，由于具有很强的数据表征学习能力以及复杂的模式识别能力，给传统检测技术的局限性提供了一种新的解决思路。本文主要研究怎样把深度学习方法系统地应用到网络入侵检测领域，利用特征学习过程、分类模型结构和系统处理流程三者相互配合来创建一个更加高效、精确、智能的检测系统。本文的研究工作，一方面是对已有技术的改良尝试，另一方面也是对今后智能化安全防御体系搭建的一次探查。

1 网络入侵检测技术的演进与挑战

1.1 传统检测范式的固有局限

传统的入侵检测技术主要是基于误用检测和异常检测两种范式，但是它们的内在机制存在着深刻的不足。误用检测很大程度上依赖于事先建立好的攻击特征库，就像通过“通缉名单”进行比对一样。该模式不能识别不在特征库内的新型或者变种攻击（零日漏洞），容易受到混淆、加密等技术的绕过。异常检测则试图用建立正常行为基准的方式来发现异常，但是如何界定正常行为的范围很难，容易造成两种错误：合法业务波动被误判为攻击（高误报），或者不能识别出模仿正常行为的高级威胁（高漏报）。除此之外，网络规模不断扩大，流量加密越来越普遍，攻击方式越来越协同化，传统的处理方法在实时性、扩展性以及对加密内容的解析能力上都遇到了很大的困难。

1.2 人工智能驱动的技术革新优势

人工智能，尤其是机器学习、深度学习给入侵检测领域带来了新的可能性。不同于依靠人工先验知识，它们可以从大量的历史网络流量数据中，利用算法自动学习并提取出区分正常和恶意行为的深层次、非线性特征模式[1]。深度学习模型，如卷积神经网络，可以自动地从原始数据包字节流到高级攻击语义生成层次化特征表示，不需要复杂的特征工程。循环神经网络以及它的变体 LSTM 非常适合分析网络连接、会话序列里面蕴含的时序依赖关系，因此可以很好地检测 DDoS 攻击、横向移动这些带有时间关联特征的威胁。生成对抗网络等新技术可以用来合成攻击样本以增强模型训练，也可以模拟对抗性环境来提高模型的鲁棒性。更重要的是具有在线学习能力的 AI 模型可以不断适应网络环境、攻击手法的变化，使检测策略不断进化，从根本上解决了传统系统适应性差的问题。

2 基于深度学习的入侵检测系统架构设计

2.1 层次化智能特征提取框架

本研究给出了一种多层次、自适应的智能特征提取框架。该框架采用端到端学习方式，直接从预处理网络流量中学习特征表达。第一层一般用一维卷积神经网络，用来从原始数据包或者网络流中抓取局部基础模式。第二层使用循环神经网络单元来分析跨数据包或者流的时序关联规律，从而发现攻击行为在时间维度上演变的特点。为了提升特征提取的精确度和效率，框架中嵌入

了注意力机制。该机制可以给不同的时序或者特征通道赋予不同的权重，使模型集中于对攻击识别贡献较大的关键信息，抑制冗余或无关干扰，从而生成判别性较强的综合特征向量。

2.2 优化的深度神经网络分类器构建

获得高效特征之后，建立稳健的分类器是实现精确检测的基础。本文采用并改进了深度残差网络为基本模型。其跳跃连接机制很好地缓解了深层网络训练时的梯度问题，保证了模型有足够深的表达层次。针对网络攻击数据中类别分布不均衡的问题，设计了加权交叉熵损失函数，给少数类样本赋予更大的损失权重，加强模型对稀有攻击模式的学习，减小类别不平衡造成的模型偏差。另外使用了集成学习策略，在不同的数据子集或者特征视角上训练多个基分类器，并用软投票、堆叠泛化等方法把各个基分类器的决策结果结合起来，降低单个模型的决策方差，提高整体检测结果的准确性和稳定性。

2.3 面向实时的系统性能增强策略

为了满足网络环境对实时性的严格要求，采取了各种系统性能优化策略。利用知识蒸馏等模型压缩加速技术，将教师模型的知识迁移到轻量化学生模型上，大幅度降低模型的计算复杂度、存储消耗，同时保持和原始模型相当的高检测精度。第二，设计 GPU 加速的高效并行处理流水线，对数据捕获、特征提取、分类推断等过程进行并行化、流水线处理，获得最佳系统吞吐量。第三，增量学习，支持模型在线上不断使用新的数据进行学习，从而达到不断进化检测的效果。最后建立动态阈值调节机制，使系统可以按实时安全态势来自动调节告警触发条件，实现安全运维管理精细化、智能化。

3 实验设计与模型评估方法论

3.1 实验环境与数据准备

为了对提出的系统进行全面的性能评测，在如上实验平台的基础上建立了一个接近真实的环境。该平台模拟了一个中型企业网络，其中有各种类型的服务器、工作站、网络设备，可以产生背景流量和模拟攻击流量。在数据集上使用 NSL-KDD、CICIDS2017 等公认的基准数据集。数据集中包含从正常网络行为到拒绝服务、端口扫描、暴力破解、渗透攻击等各种攻击的大样本，并且已经预先去除了冗余部分，可以用于训练以及评测基于数据驱动的检测模型[2]。实验软件框架以 Tensor Flow、PyTorch 等主流的深度学习库为基础构建，并且有高性能的 GPU 计算卡来加快模型训练过程。为了保证评估的公平、全面，我们设置了多个对比基线，即经典的基于签名的检测工具 Snort、传统的机器学习模型（随机森林、支持向量机）和没有经过优化的基础深度学习模型。

3.2 综合性能评估指标体系

衡量一个入侵检测系统好坏的标准是多维度的、严谨的评估指标。本研究主要使用的核心指标有：检测率（Detection Rate 或 Recall），即被正确识别的攻击样本在全部真实的攻击样本中所占的比例，可以用来衡量系统发现威胁的能力；误报率（False Positive Rate），指正常样本被错误判定为攻击的比例，关系着安全运维的效率和可信度；精确率（Precision），指所有的被判定成攻击的样本中有多少是真实的攻击，用来衡量模型在两类样本不均衡情况下整个系统的性能；F1 分数是精确率和召回率的调和平均数，可以综合反映系统在两类样本不均衡情况下整体的性能[3]。除了上述分类性能指标之外，系统效率指标也不可忽视，处理延迟，即系统从接收到数据到发出告警所经历的时间，反映了系统的实时性；吞吐量，即单位时间内系统可以处理的网络流量，体现了系统的可扩展性。模型的复杂度，比如参数数量、存储占用、推断能耗等，都是用来判断模型是否适合在现实环境中部署的重要依据。

3.3 实验结果分析与讨论

通过 NSL-KDD 等标准数据集的系统实验可知，本文所提出的优化深度学习检测系统，在多个重要的指标上都比传统的算法、未优化的基线深度学习模型要好[4]。系统在保持很高的检测率（识别出绝大部分攻击）的同时，将误报率控制在一个很低的水平上，实现了精确率和召回率的良好平衡，F1 分数达到领先水平。特别值得一提的是，对那些传统方法难以检测的、复杂度高或者样本量少的攻击类型（U2R、R2L 类），本系统具有更好的表现。在实时性方面，优化后的系统通过使用轻量化模型、并行流水线实现了微秒级单样本的处理延迟及高吞吐量，满足高速骨干网络对于实时监控的需求。采用消融实验的方法，对注意力机制、残差结构、类别平衡损失等关键组件对最终性能提升的贡献度进行定量分析，验证各个设计模块的有效性和必要性。模型压缩技术的使用使得深度模型可以部署在资源有限的边缘设备上。

4 系统鲁棒性与安全性考量

4.1 对抗性攻击的防御策略

人工智能模型本身可以被攻击者攻击，受对抗性样本的攻击。攻击者可以构造出人眼几乎察觉不到的扰动输入，使模型出现错误的分类。入侵检测当中，这就表明攻击者可以对恶意流量进行微调，让其伪装成正常流量，在特征空间里逃避 AI 检测模型[5]。因此，必须把系统的鲁棒性考虑进去。本研究探索的途径有，在训练时加入对抗样本，也就是主动地生成对抗样本来加入到训练集中，使模型学会辨别和抵御这些对抗；对输入特征进行标准化和平滑处理，以降低模型对于小扰动的敏感性；利用集成方法的多样性，不同的基分类器对同一对抗样本可能会表现出不同的脆弱性，集成的投票可以提高整体的鲁棒性。

4.2 模型可解释性与决策信任

深度学习模型经常被诟病为“黑箱”，内部的决策过程无法理解。缺少可解释性会影响安全分析师对告警的信任度以及应急响应的效率。因此提高模型的可解释性就成为一项重要工作。我们试着用 LIME、SHAP 这样的事后解释方法，给单个样本的预测结果做特征贡献度分析，清楚地显示是哪些网络流量特征（比如某个特定端口的连接次数，数据包长度的异常情况等）引发了“攻击”的判定。另外探索可解释性更强的模型结构（比如注意力权重的可视化）或者设计规则和神经网络相结合的混合模型也是提高决策透明度、形成人机协同的安全分析闭环的努力方向。

5 实际部署与集成应用探讨

5.1 云端与边缘协同部署模式

在网络环境中，流量的大小以及计算资源的分布是不均匀的，单一的部署模式不能满足所有的场景。我们研究一种云-边协同的智能检测架构。在网络边缘，比如分支机构、IoT 网关处，部署经过极大压缩和优化的轻量级检测模型，对数据进行初步的、低延迟的本地过滤与异常感知，把可疑流量或者元数据摘要上报到云端安全中心。云端汇聚各边缘节点的信息，运行更复杂的、更强大的检测模型，对信息做深度关联分析以及全局威胁情报研判。该模式既可以利用边缘计算的实时性、数据本地化处理的隐私性优势，也可以发挥云端强大的计算和数据分析能力，达到纵深防御的目的[6]。

5.2 与现有安全体系的集成

一个好的 AI 入侵检测系统不应该孤立存在，而应该和组织内部现有的安全设备和流程无缝集成。本研究讨论了同安全信息与事件管理系统之间的整合方案，使得 AI 模型发出的告警可以同防火墙、终端检测响应等其它安全日志实现关联分析，从而得到更为完整的攻击链视图，并且可以联动网络策略执行点，比如当检测到高置信度的攻击时，通过 API 自动下发指令给下一代防火墙或者交换机，临时阻断恶意 IP 或者会话，从检测到响应的自动化闭环中提升主动防御能力。

6 结论与展望

6.1 结论

本文系统地研究把深度学习技术应用到网络入侵检测系统中，从特征提取、模型构建、性能优化、鲁棒性增强等多方面给出解决措施。经过实验，得到的设计方法可以提高检测的精度，降低误报率，满足实际部署时对于实时性和效率的要求，给日益复杂的网络威胁提供有力的智能化工具。但是这一领域仍然存在很多挑战并且有广阔的研究空间。

6.2 展望

未来的研究可以关注以下前沿方向，多模态异构数据融合发展，形成全方位的威胁感知；探索隐私计算框架，实现安全前提下的协同模型训练和知识共享；加强 AI 模型本身的抗攻击鲁棒性研究；轻量化、可解释模型在 5G、物联网等复杂网络环境中的部署与应用。利用持续技术融合和创新去建设下一代的智能安全防御系统。

参考文献

[1]张杨. 人工智能时代网络入侵检测与防御技术[J]. 数字通信世界,2025,(07):71-73.
[2]路博. 人工智能赋能下的网络入侵检测精准模型构建[J]. 中国宽带,2025,21(08):34-36.
[3]王斌. 人工智能在网络入侵检测系统中的应用与优化策略[J]. 中国宽带,2025,21(03):61-63.
[4]王俊恒. 人工智能与自适应算法在网络入侵检测与防御策略中的应用[J]. 集成电路应用,2025,42(01):262-263.
[5]刘广睿. 智能网络入侵检测系统的安全性评估与防御技术研究[D]. 哈尔滨工业大学,2024.
[6]刘恩锴. 基于人工智能的网络入侵检测与防御技术研究[J]. 网络安全技术与应用,2024,(08):26-28.