基于改进人工免疫的计算机网络异常数据流入侵检测方法

刘 彧[1]
LIU Yu

摘 要 在应对当前计算机网络中大规模、动态、时变的实时数据流时，传统入侵检测方法由于采用固定时间间隔的检测方式，无法实时捕捉数据流中的动态变化，导致当异常行为发生时，检测模型仍在处理旧数据，不能及时响应，从而造成检测结果滞后、命中率低等问题。为此，文章提出一种基于改进人工免疫的计算机网络异常数据流入侵检测方法。该方法对计算机网络中的数据流进行预处理，以消除噪声并增强数据的可分析性。从预处理后的数据流中精准提取出能够表征异常行为的特征。基于这些特征，建立一种在改进人工免疫算法框架下的异常数据流入侵检测模型。该模型通过引入改进的免疫机制，增强算法对复杂网络环境的适应性和鲁棒性。同时，模型采用滑动时间窗口技术对实时数据流进行连续的增量更新，确保检测模型始终基于最新的数据流信息进行异常检测，以此提高检测的时效性和准确性。实验结果表明，该方法检测命中率超过 92%，显著优于其他方法，充分证明了该方法在提升检测敏感度和特异性方面的有效性。此外，该方法还有效降低了误报和漏报，为无线网络安全提供了坚实的技术保障。

关键词 改进人工免疫；计算机网络；异常数据流；异常数据流入侵；入侵检测

doi：10.3969/j.issn.1672-9528.2025.01.008

0 引言

网络环境的日益复杂带来了前所未有的安全挑战，其中异常数据流的入侵检测成为网络安全领域的研究热点。传统的入侵检测方法在面对日益多样化的网络攻击时，往往存在检测率低、误报率高、适应性差等问题，难以满足当前网络安全的需求。在此背景下，学者研究出多种检测方法。其中基于最大频繁项的数据流异常检测，通过挖掘数据流中的频繁项集，分析这些项集在时间序列中的变化规律，从而识别出异常行为[1]。这种方法在大规模数据流中，频繁项集的挖掘是一个计算密集型任务，特别是当数据流中的项数量和数据流的速度都非常大时。挖掘频繁项集需要较长时间，导致系统无法实时处理数据流中的新数据，造成检测结果滞后，难以达到理想的检测效果。基于生成对抗网络的计算机异常数据入侵检测分析通过生成器和判别器的对抗训练，学习数据的内在分布特征，进而识别出与正常模式显著偏离的异常数据。这种方法在理论上具有较高的检测精度和泛化能力，能够应对新型和未知的网络攻击。然而，GAN 模型依赖于大量的训练数据来学习数据的内在分布特征，并且训练过程涉及多次迭代和参数调整，因此整个训练周期相对较长。当数据流中的特征发生显著变化时，已经训练好的 GAN 模型可能无法立即适应这些变化，从而导致检测性能下降和无法及时响应的问题。因此，为解决传统方法检测滞后和命中率低的问题，提出一种基于改进人工免疫的计算机网络异常数据流入侵检测方法。

1 计算机网络数据流预处理

在计算机网络环境中，海量的数据流持续不断地在网络中传输，这些数据流中既包含了正常的网络活动信息，也可能隐藏着异常或恶意的入侵行为。因此，需要对这些海量的数据流进行预处理[3]。

先对数据进行清洗，设 D_{raw} 为原始数据流集合，D_{noise} 为噪声和冗余信息集合，则数据清洗函数 f_{clean} 可以表示为：

$$D_{cleaned}=f_{clean}(D_{raw}, D_{noise}) \quad (1)$$

$D_{cleaned}$ 实际上是一个动态定义的集合，依赖于具体的清洗规则、正则表达式等技术手段。

接着进行数据解析，针对 HTTP 数据流解析 HTTP 头部信息、URL 参数等了解请求细节；对于数据库日志则关注 SQL 语句执行情况、访问权限变化等关键信息[4]。

最后，经处理的数据将被送入后续的分析和决策支持系统，用于发现潜在安全威胁、优化网络性能和提升用户体验等。

2 提取数据流中异常行为特征

完成数据流与处理后，提取数据流中异常行为特征，为

1. 陕西交通职业技术学院 陕西西安 710018

了实现这一目标，引入 MOA 这一强大的数据流挖掘平台。根据 MOA 平台提供的算法库，选择基于聚类的异常数据流特征提取算法。在明确哪些特征值或特征组合被视为异常之前，需要深入理解网络流量的正常行为模式[5]。可以定义以下分类范畴作为异常检测的基准：

（1）异常流量模式：如突发的流量峰值、非周期性的流量波动、长时间持续的异常低流量等。

（2）非标准协议使用：检测到非标准端口上的常见协议活动，或标准端口上的非标准协议活动。

（3）源 / 目标地址异常：来自或发往已知恶意 IP 地址的流量，或来自未知 / 罕见地理位置的流量。

（4）会话异常：异常短的会话（表示扫描活动）或异常长的会话（表示数据泄露）。

构建 k 个聚类中心 $A_j(L)$，其中 j 遍历 1 至 k，以此为基础对网络流量进行聚类分析。在聚类过程中，计算每个聚类簇的根节点与簇心之间的距离是至关重要的一步，它有助于更清晰地划分数据边界，识别出潜在的异常行为特征[6]。计算表达式为：

$$|x| = d_j(x, A_j) = \sqrt{\sum_{l=1}^{|L|}(x_l - A_j(l))^2} \tag{2}$$

式中：$|L|$ 是异常数据流特征空间 L 的维度；x_l 和 $A_j(l)$ 分别是点 x 和聚类中心 A_j 在特征 L 上的值。

3 建立改进人工免疫算法的异常数据流入侵检测模型

基于上述章节中提取出的这些特征，进一步构建基于改进人工免疫算法的异常数据流入侵检测模型。该模型借鉴生物免疫系统的原理和机制，通过模拟免疫细胞的识别、学习和记忆过程，能够自动学习并更新异常行为特征库，实现智能化的异常检测。这有助于提高检测效率，实现对网络异常行为的自动检测和响应。

改进人工免疫算法（improved artificial immune algorithm，IAIA）是在传统人工免疫算法（artificial immune algorithm，AIA）的基础上，通过引入新的策略、优化算法参数或结合其他算法思想，以提高其性能、收敛速度和求解精度的一种算法[7]。改进人工免疫网络入侵检测模型的环境配置表如表 1 所示。

表 1 改进人工免疫网络入侵检测模型环境设置表

检测模型指标名称	初始值	边缘值（最小值 / 最大值）
检测密度比	0.5	0.01/0.99
迭代次数 / 次	100	10/500
映射值	0.75	0.0/1.0
单元攻击频次 / 次	50	1/1000
检测率 DR	90%	50%/99.9%

在此基础上，构建高效、自适应的入侵检测架构，该架构以实际需求为导向，在每个检测层级设定目标与执行策略。此架构深度融合改进人工免疫学的核心理念，形成一个结构清晰、功能全面的检测体系框架，如图 1 所示。

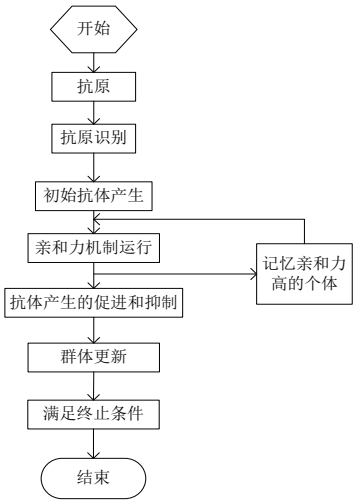

图 1 人工免疫系统的通用框架

在架构的初始化阶段，为精确评估网络威胁，引入了基于信息层与物理层双重维度的亲和度评估机制，该机制通过目标函数乘积量化潜在攻击路径（抗原）的破坏力。同时，构建了检测器系统，每个检测器由中心向量 $\boldsymbol{c}$ 与检测半径 d 定义，共同界定一个超球体检测空间。一旦数据点落入此空间，即判定为非自体元素，立即触发防御机制。检测逻辑：通过计算抗原特征向量 $|\boldsymbol{x}|$ 与检测器中心向量 $\boldsymbol{c}$ 的欧几里得距离 d_E，并与检测半径 d 对比，若距离小于等于检测半径，则判定抗原为非自体，随即启动防御措施：

$$d_E(|x|, c) = \sqrt{\sum_{i=1}^{n}(|x_i| - c_i)^2} \tag{3}$$

如果 $d_E(\boldsymbol{x}, \boldsymbol{c}) \le d$，则抗原 $\boldsymbol{x}$ 被视为非自体，触发防御机制[8]。

此基于改进人工免疫算法的异常数据流入侵检测模型，不仅能够显著增强网络系统抵御复杂网络攻击的能力，还展现出强大的自学习和自适应特性。并根据网络环境及攻击手法的演变，自动调整优化检测策略，实现对新兴网络威胁的快速识别与有效防御。

4 数据流增量更新输出检测结果

在实际应用中，网络数据流是实时变化的，因此入侵检测系统需要具备对数据流进行增量更新的能力。当新的数据流到达时，系统能够迅速对其进行处理、特征提取，并输入到已建立的检测模型中进行实时检测。通过引入滑动时间窗口技术等手段，改进人工免疫算法能够对实时数据流进行连续的增量更新，确保检测模型始终基于最新的数据流信息进行异常检测。这有助于及时发现并响应新的网络威胁，提高检测的时效性。最终，系统会将检测到的异常数据流及其相关信息输出，以便网络管理员或安全系统采取相应的防御措施，保护网络免受攻击和侵害。

图 2 直观展示了网络入侵实时检测框架的运作流程，从数据流的接收、特征提取，到模型的实时检测与动态调整，再到结果的即时输出，各环节紧密衔接，构建起一个强健、敏捷的网络安全防护体系 [9]。在此框架下，数据流在网络中的流动与监控变得透明，系统能迅速识别并应对潜在威胁。

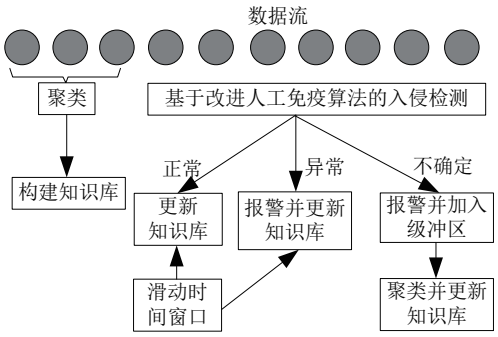

图 2 网络入侵实时检测框架

滑动时间窗口（SW）作为核心组件，如同智能筛网，高效筛选实时数据，剔除过时信息。其大小依窗口持续时间和数据流速率而定，初始化时设定大小与滑动步长。

（1）窗口初始化：设定窗口大小 W（固定数量的数据项），以及窗口的滑动步长 S（时间间隔）。

（2）数据接收：对于每个新到达的数据项 d_t（在时间 t 到达），将其加入窗口。用表达式可表示为：

$$SW \leftarrow SW \cup \{d_t\} \tag{4}$$

（3）数据剔除：当数据项 d_{t-T}（在时间 t~T 到达，且 T 为窗口持续时间）超出窗口范围时，将其从窗口中移除，整体过程可描述为：

$$SW \leftarrow SW - \{d_{t'} \mid t - t' > T\} \tag{5}$$

面对数据流的高速更新，算法展现极高敏捷性，即时调整滑动窗口树，确保与数据流同步，捕捉动态变化，保障分析结果的时效与准确。同时，数据流演变驱动分析模型进化，新训练记录触发 Max P-tree 及其支持度的重新评估，整合至 Max FP-Tree 更新流程中，实现对数据准备评价。最终输出的 Max FP-Tree，既是数据流最新检测结果的实际记录，也是关键模式信息的深度挖掘，为数据驱动决策奠定坚实基础 [10]。

5 实验

5.1 实验准备

本次实验采用最新的 CICIDS2017 数据集作为评估入侵检测模型效能的基准。该数据集由某网络安全研究所构建，旨在全面反映当前网络环境中复杂的攻击模式与正常流量模式。数据集涵盖了为期五天的网络流量，不仅包含了常见的网络攻击类型（如 DoS 攻击、Brute-Force、Web 攻击等），还融入了最新的网络威胁，如物联网（IoT）攻击和高级持续性威胁（APT）模拟场景，以充分考验检测系统的时效性和准确性方面的表现。

为了确保实验结果的可靠性和可重复性，选择了符合实验具体需求、数据处理的能力要求以及成本效益平衡的软硬件环境。这一选择综合考量了多方面因素，旨在为实验提供一个稳定且高效的基础平台。本次实验所采用的详细软硬件配置如表 2 所示。

表 2 实验软硬件选择

设备 / 软件	型号 / 版本	使用数量
服务器	Dell PowerEdge R740xd	1 台
CPU	Intel Xeon Gold 6248R (2.7 GHz, 24 核)	2 颗
内存	DDR4 ECC 256 GB	8 条
存储	Samsung PM981a 1 TB NVMe SSD	2 块
网络接口卡	Intel X710-DA4 4x10GbE	1 张
操作系统	Ubuntu Server 20.04 LTS	—
编程语言	Python 3.8	—
流量生成器	Spirent TestCenter	1 套

5.2 实验数据与指标

本研究基于 CICIDS2017 数据集的变体，构建了包含 60 万条训练记录和 1.5 万条独立测试记录的数据集，以评估入侵检测模型的泛化能力和精度。测试集多样性丰富，包含 68.5% 正常流量、20.1% 明确标记的异常入侵和 11.4% 未知异常入侵。实验核心评估指标为检测命中率，即系统正确识别攻击事件的比例。通过训练集构建模型，测试集评估其性能，对比系统判定与真实标签，统计正确识别的攻击数量。高检测命中率表明系统性能优越，能有效识别网络攻击，保障网络安全。

5.3 实验结果与分析

本次实验旨在验证在入侵检测模型中引入滑动时间窗口技术后，对实时数据流进行连续的增量更新能否显著提升检测模型的时效性，减少检测滞后性。在 Linux 操作系统，Python 编程环境下，基于上述训练集构建基于改进人工免疫算法的入侵检测模型，但不使用滑动时间窗口技术。在相同算法框架下，引入滑动时间窗口技术。窗口大小根据实验需求设定，如每 5 s 更新一次数据。模型持续接收新数据，并基于最新的窗口数据进行检测。确保测试集包含连续、实时的网络流量数据。使用未引入滑动时间窗口的模型对测试集进行检测，记录检测滞后时间（即检测到异常与异常实际发生之间的时间差）。随后，使用引入滑动时间窗口的模型对同一测试集进行检测，同样记录相关指标。使用与未使用滑动时间窗口技术的入侵检测模型的检测滞后性对比结果如表 3 所示。

表 3 应用滑动时间窗口技术前后入侵检测模型检测滞后性对比

检测方法 / 指标	未应用滑动时间窗口技术	应用滑动时间窗口技术（窗口大小：5 s）
平均检测滞后时间	0.83 s	0.02 s
最大检测滞后时间	3.20 s	0.50 s
最小检测滞后时间	0.10 s	0.01 s
处理速度（每秒数据量）	43.8 GB/s	84.7 GB/s
模型更新频率	批量更新	每 5 s 增量更新

分析表 3 可以看出，使用滑动时间窗口技术后，平均检测滞后时间从 0.83 s 显著降低到 0.02 s，这表明模型在检测到异常时的响应时间大大加快，从而显著提升了检测的时效性。最大检测滞后时间从 3.20 s 减少到 0.50 s，虽然仍然受到一些外部因素的影响，但整体上表明了模型在处理极端情况下的响应能力也有所增强。这有助于在紧急情况下更快地发现并响应潜在的威胁。处理速度从 43.8 GB/s 提升至 84.7 GB/s，表明模型在单位时间内能够处理更多的数据，这对于处理实时数据流至关重要。更快的处理速度意味着模型能够更快地分析新数据，从而更快地检测到潜在的异常。此外，模型更新频率从批量更新（如每 10 min）变为每 5 s 增量更新，这是滑动时间窗口技术的直接效果。通过持续不断地接收和处理新数据，模型能够更紧密地跟踪数据流中的变化，从而增强了检测的实时性。通过对比实验，证明了在入侵检测模型中引入滑动时间窗口技术能够显著提升检测模型的时效性，减少检测滞后性，从而更有效地应对实时网络流量中的潜在威胁。

为了进一步验证所提方法的综合有效性，本次实验对比了基于最大频繁项的数据流异常检测（方法 1）、基于生成对抗网络的计算机异常数据入侵检测分析（方法 2）以及本文提出的创新方法。不同方法在无线网络入侵检测任务中的表现如图 3 所示。

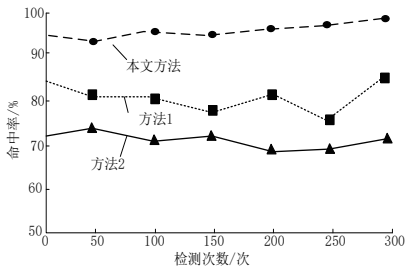

图 3 三种方法的命中率对比

在全面对比本文提出的入侵检测方法和另外两种方法后，实验结果（如图 3 所示）清晰地展现了本文方法的显著优势。本文方法实现了超过 92% 的稳定命中率，这一显著优势不仅远远超越了方法 1 和方法 2 的命中率上限（两者均未触及 85% 的门槛），更凸显了其在应对复杂、动态网络环境中的高度适应性和准确性。本文方法的核心在于构建了一个基于改进人工免疫算法的异常数据流入侵检测模型。该模型通过引入增强的免疫机制，显著提升了算法在复杂网络环境中的适应性和鲁棒性。这种机制使得模型能够更加灵活地应对各种未知和变种的攻击行为，从而提高了检测精度。

6 结语

本文所提出的基于改进人工免疫的计算机网络异常数据流入侵检测方法，在显著提升检测效能方面取得了显著成效。该方法的检测命中率高达 92%，并有效降低了误报和漏报，为网络安全防护提供了有力支持。然而，面对网络空间日益增长的复杂性和攻击规模，其性能边界仍需不断探索与突破。特别是在处理极端网络环境下的高速攻击流时，如何进一步优化资源分配、提升处理速度，将是后续研究的重点。

参考文献：

[1] 史晓晨 . 基于最大频繁项的数据流异常检测 [J]. 电脑知识与技术 , 2022, 18(25): 118-120+125.

[2] 朱昱林 . 基于生成对抗网络的计算机异常数据入侵检测分析 [J]. 集成电路应用 , 2022, 39(8): 124-125.

[3] 艾洲 . 基于改进人工免疫算法的电力系统网络攻击破坏程度量化研究与应用 [J]. 广西电力 , 2023, 46(5): 53-61.

[4] 彭国勇 . 基于神经网络的企业内部网络恶意入侵检测方法 [J]. 信息与电脑 (理论版), 2023, 35(19): 65-67.

[5] 苏新 , 田天 , 周一青 , 等 . 基于异常行为的海洋气象传感网的入侵检测方法研究 [J]. 通信学报 , 2023, 44(7): 86-99.

[6] 王国仕 , 覃岩岩 , 王初平 , 等 . 基于人工免疫的高效通信网络入侵检测方法 [J]. 长江信息通信 , 2023, 36(7): 201-203.

[7] 张海霞 . 基于网络爬虫技术的校园网入侵信息跟踪研究 [J]. 三门峡职业技术学院学报 , 2022, 21(4):135-140.

[8] 王子杰 , 潘啸天 . 工业变电站运维系统异常数据入侵检测互信息实现 [J]. 现代工业经济和信息化 , 2024, 14(8): 94-95+104.

[9] 江荣 , 刘海天 , 刘聪 . 基于集成学习的无监督网络入侵检测方法 [J]. 信息网络安全 , 2024, 24(3): 411-426.

[10] 陈海文 , 余员琴 , 王叶 , 等 . 基于概念漂移的集成增量学习 WSN 入侵检测方法研究 [J]. 网络安全技术与应用 , 2022(8): 29-32.

【作者简介】

刘彧（1989—），男，陕西西安人，硕士，讲师，研究方向：计算机应用技术。

（收稿日期：2024-09-24）