关联规则下计算机网络多阶段攻击异常流量检测方法

孙月玲 [1]
SUN Yueling

摘 要 多阶段攻击包含侦察、渗透、横向移动等多个相互关联的阶段，各阶段行为具有内在的时序关联性和逻辑依赖关系。传统的检测方法缺乏对攻击全流程的系统性关联分析，无法构建完整的攻击链，难以全面、准确地识别多阶段攻击的整体模式和意图，导致对复杂多阶段攻击的检测能力受限。为此，文章提出基于关联规则的异常流量检测方法。利用 NetFlowMeter 工具提取原始流量的 85 维特征形成结构化矩阵，过滤无效样本后采用基于方差阈值的特征选择算法（featureselection-variancethreshold, FS-VT）降维，并将数值特征转换为灰度图像，完成流量数据预处理。通过模糊集合处理连续特征并采用自适应支持度阈值，结合改进关联规则算法（Apriori）挖掘网络异常流量强关联规则。针对多阶段攻击特性，设计多阶段攻击关联图模型，通过定义各阶段规则集并计算规则相似度，结合混合残差网络提取的深度特征进行加权融合，依据综合评分实施分级响应。实验结果表明，该方法检测准确率达 98.7%，误报率为 0.45%，漏报率为 1.3%，且各攻击阶段检测指标均表现优异，能够有效应对复杂网络环境中的高级持续性威胁。

关键词 关联规则；计算机网络；多阶段；攻击；异常流量

doi：10.3969/j.issn.1672-9528.2026.01.026

0 引言

移动互联网的普及，使得网络安全问题日益突出。当前，针对互联网的攻击方式已从传统病毒、木马发展到现在的“高级持续性威胁”“零天漏洞”等多种攻击方式，已成为威胁用户隐私及重要基础设施安全的重要威胁。移动互联时代，网络的多样性、开放性和终端的智能性进一步加剧了网络安全的复杂性。面对海量、高维、动态变化的网络业务，传统的基于规则或特征的检测方法存在较大的滞后性和虚警率，很难有效应对多阶段、高隐蔽性的网络攻击 [1]。如何高效、准确地发现异常流，特别是多阶段攻击，是目前网络安全面临的一个重要问题。

多阶段攻击是一种多个相互联系的攻击，在最初的几个阶段中，攻击者只会进行一些简单的试探，或者只是简单地收集一些信息，然后逐渐深入，最后造成数据盗窃或系统瘫痪。在此基础上，通过对多个节点的相互关系进行全面分析，从而实现了对单个节点的检测。关联规则分析是一种典型的数据挖掘方法，可以从海量的网络业务数据中挖掘出潜在的攻击模式，并构建攻击各阶段之间的相关性，进而实现对多个阶段的攻击行为的综合检测 [2]。但是，对于高维、动态变化的网络业务，传统的关联规则算法面临着运算复杂度高、自适应性差等问题。近年来，随着机器学习方法的飞速发展，该问题的研究成为一个新的研究方向。在此基础上，将机器学习在数据挖掘方面的优势与关联规则的良好解释性相结合，建立鲁棒、智能化的网络异常检测模型。基于此，本文将开展关联规则下计算机网络多阶段攻击异常流量检测方法研究。

1 基于特征工程与图像转换的网络流量预处理

原始计算机网络流量数据（如 pcap 格式）普遍存在高维度、异构性和噪声干扰等问题。高维度导致数据包含大量冗余或不相关特征，异构性表现为数据来源多样且结构不一致，而噪声干扰会显著降低数据质量。这些特性直接影响异常流量检测的性能，导致准确率下降和效率降低 [3]。针对这一问题，需要通过预处理将原始流量转化为标准化、高质量的数据形式。为此，本研究提出网络流量预处理方法：先借助工具提取多维特征，接着过滤无效样本，最后基于方差阈值进行特征降维以消除冗余。通过这一流程，可有效提升后续异常检测算法的准确性、效率及可靠性，进而更好地保障网络安全。

在数据表示阶段，引入改进后的 NetFlowMeter 工具，利用该工具将原始流量转化为结构化特征矩阵。假设原始流量数据集可表示为 $R = (r_1, r_2, \cdots, r_m)$，其中 r_i 表示第 i 个流量样本。通过对数据特征的提取，每个样本都能够被映射为 d 维特征矢量 $\boldsymbol{x}_i = (\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \cdots, \boldsymbol{x}_i^d)$，最终得到标准的数据集 $\boldsymbol{X}_0 = \{x_1,$

1. 江苏联合职业技术学院 江苏盐城 224000

$x_2,\cdots, x_n\}$。通常情况下，当特征维度 d 取值为 85 时，涵盖流量统计特征（如包长度均值、流量速率）、协议特征（如 TCP/UDP 比例）和行为特征（如连接持续时间）。

数据的过滤阶段，其目标是将无效样本剔除。对无效样本的定义包括：包含缺失值（NaN）的样本；包含无穷大值（±inf）的样本；特征值超出合理范围（如负的包长度）[4]。通过过滤操作，得到净化后的数据集 $\boldsymbol{X}_f \subseteq \boldsymbol{X}_0$。

在特征降维阶段，采用基于方差阈值的 FS-VT 算法（Feature Selection-Variance Threshold）。该算法的核心是保留方差大于阈值的特征，剔除低方差特征，以此降低数据的冗余。这一过程中每个特征的方差计算用公式表示为：

$$\mathrm{var}(\boldsymbol{x}^j)=\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{x}_i^j-\bar{\boldsymbol{x}}^j\right)^2 \tag{1}$$

式中：$\boldsymbol{x}^j$ 表示第 j 维特征量；$\bar{\boldsymbol{x}}^j$表示第 j 维特征的均值。在此基础上，选择方差在区间 [0,1] 内具有最大间隙值的方差作为阈值。过滤方差中低于阈值的特征量，生成降维后的数据集[5]。为进一步提高后续异常流量检测的精度，将数值型的特征转化为灰度图像。对上述生成的数据集进行归一化处理，其表达式为：

$$\boldsymbol{x}_i^j=\frac{\boldsymbol{x}_i^j-\min(\boldsymbol{x}_i)}{\max(\boldsymbol{x}_i)-\min(\boldsymbol{x}_i)}\times 255 \tag{2}$$

在此基础上，对每个样本 $\boldsymbol{x}_i$ 进行重塑，得到 $w\times h$ 的灰度图像，其中 w 和 h 的取值均由特征维度决定。例如，当特征数为 64 时，此时可将其转化为 8×8 的图像矩阵，最终完成对图像数据集的生成。

2 网络异常流量关联规则挖掘

在计算机网络多阶段攻击异常流量检测过程中，完成网络流量预处理后，虽然得到了标准化、高质量的数据，但这些数据中潜在的攻击模式和关联关系仍未被充分揭示。网络攻击往往具有复杂的模式和多个阶段的特征，仅依靠预处理后的数据难以直接识别出这些隐藏的攻击行为[6]。因此，需要进行网络异常流量关联规则挖掘，从预处理后的数据中深入分析流量特征之间的内在联系，以发现攻击的高频率模式。本文采用改进的 FARM 方法，将模糊集理论与自适应支撑策略相结合，有效挖掘异常业务的关联规则。

网络流量特征通常呈现为连续值，例如流量速率、包数量等。若对这些连续值特征进行离散化处理，会产生尖锐边界问题。为避免这一问题，将每个特征划分为模糊集合[7]。以 TCP 包比例为例，定义其模糊集合为 {low, medium, high}，其隶属度函数可表示为：

$$\mu_{\mathrm{low}}(\boldsymbol{x})=\begin{cases}1, \boldsymbol{x}\le 80\\ -0.1\boldsymbol{x}+9, 80<\boldsymbol{x}<90\\ 0, \boldsymbol{x}\ge 90\end{cases} \tag{3}$$

$$\mu_{\mathrm{high}}(\boldsymbol{x})=\begin{cases}0, \boldsymbol{x}\le 85\\ 0.15\boldsymbol{x}-12.75, 85<\boldsymbol{x}<95\\ 1, \boldsymbol{x}\ge 95\end{cases} \tag{4}$$

考虑到传统固定支持度阈值难以适应计算机网络这种动态变化的流量环境。因此，结合二项分布，提出自适应支持度阈值。计算频繁项集数量的概率分布，其公式为：

$$P(a)=C_n^a p^a(1-p)^{n-a} \tag{5}$$

式中：a 表示频繁项集数量；p 表示项集成为频繁项集的先验概率。再通过简单计算得到平均支持度，利用数据库扫描次数因子对阈值进行动态调整。

利用改进后的 Apriori 算法对强关联规则进行挖掘：

（1）连接：合并 k−1 项频繁集生成候选 k 项集；

（2）剪枝：删除支持度低于 T_a 或置信度低于阈值 θ 的规则；

（3）输出形如 $X\rightarrow Y$ 的规则，其中 X 和 Y 为模糊项集。

3 基于多阶段攻击关联图模型与混合残差网络的协同检测

网络异常流量关联规则挖掘虽然能够发现部分攻击相关的规则模式，但由于网络攻击具有复杂的多阶段特性（如侦察、渗透、横向移动等），各阶段行为之间存在时序关联性，仅依靠关联规则挖掘往往只能获取局部、片段化的攻击信息，难以全面刻画攻击的整体流程[8]。为此，本文提出多阶段攻击关联图模型，通过定义各阶段行为规则，并基于置信度与支持度差异计算规则相似度，结合最短路径算法构建攻击链，从而系统性描述攻击的演进过程。为进一步提升检测精度，引入混合残差网络提取流量深度特征，并与关联规则评分加权融合，形成协同检测框架。最终根据综合评分实施分级响应。

针对多阶段网络攻击，提出多阶段攻击关联图模型。模型首先定义攻击阶段的集合定义为 $S=\{s_1, s_2,\cdots, s_k\}$，在每个阶段，$s_i$ 都有与其相对应的关联规则 R_i。例如，在侦察阶段 s_1，其规则为 { 扫描行为 → 端口探测 }；在渗透阶段 s_2，其规则为 { 漏洞利用 → 权限提升 }[9]。为量化阶段间的关联性，给定两个阶段的规则集为 R_i 和 R_j，其相似度计算公式为：

$$\mathrm{sim}(R_i,R_j)=\frac{1}{\left|R_i\right|\left|R_j\right|}\sum_{R_i,R_q}\sum_{R_j}\mathrm{sim}(R_p,R_q) \tag{6}$$

式中：$\mathrm{sim}(R_p, R_q)$ 表示规则相似度，可通过置信度和支持度差异计算得到。结合计算得到的相似度，通过最短路径算法识别关键攻击路径。

为增强检测能力，本文构建了混合残差网络与多阶段攻击关联图模型的协同框架：

特征融合阶段：将混合残差网络提取的深度特征与关联规则评分进行加权融合[10]。

分级响应：根据评分结果触发分级响应：

（1）评分 ∈[0.7,1.0]：阻断 IP 并告警；

（2）评分 ∈[0.4,0.7]：限制连接速率；

（3）评分 < 0.4：记录日志。

4 实例应用分析

4.1 应用环境

在上述关联规则下，提出一种全新的计算机网络多阶段攻击异常流量检测方法。为验证该检测方法的实际应用有效性，将其应用到真实的计算机网络环境中，对多阶段攻击异常流量进行检测。应用环境由三个主要部分组成，分别为网络流量采集层、数据处理层和检测分析层。图 1 为该测试环境网络拓扑结构图。

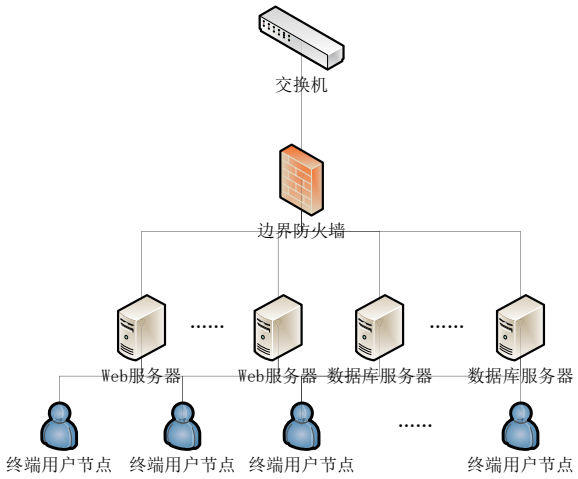

图 1 测试环境网络拓扑结构图

如图 1 所示，该环境包含核心交换机、边界防火墙、Web 服务器集群（10 台）、数据库服务器（3 台）以及 200 个终端用户节点。测试期间网络平均吞吐量为 1.2 Gbit/s，峰值流量达到 2.8 Gbit/s。

采用真实的计算机网络环境和模拟攻击相结合的方式，完成对测试环境的构建，其中正常流量来自日常业务系统运行产生的真实流量，包括 HTTP/HTTPS 访问、数据库查询、文件传输等常规网络活动。异常流量是通过 Metasploit 框架模拟生成，覆盖了典型的多阶段攻击。表 1 记录了多阶段攻击具体流程。

表 1 多阶段攻击具体流程

阶段	内容	攻击行为
第一阶段	侦察阶段	网络扫描（Nmap）、服务探测（OpenVAS）
第二阶段	初始入侵	漏洞利用（CVE-2023-1234）
第三阶段	横向移动	凭证窃取（Mimikatz）、权限提升
第四阶段	目标攻击	数据窃取（定制化恶意软件）

将测试数据集的时间跨度设置为 30 天，共捕获到 1.2 TB 的原始流量数据。在对其进行预处理后，得到结构性特征数据集。表 2 记录了测试数据集的基本组成。

表 2 测试数据集的基本组成

流量类型	样本数量	占比 /%	攻击阶段
正常流量	3 856 742	78.2	—
扫描流量	412 587	8.4	侦查
漏洞利用	185 326	3.8	初始入侵
横向移动	256 914	5.2	横向扩展
数据窃取	219 431	4.4	目标达成

按照上述内容完成对测试环境的构建。

4.2 参数配置

在应用本文上述检测方法对该测试环境中多阶段攻击异常流量检测时，按照表 3 中的数据，完成对关联规则挖掘的设置。

表 3 关联规则挖掘设置

参数	值	说明
最小支持度	0.15	自适应计算基准值
最小置信度	0.65	规则筛选阈值
滑动窗口	5 min	时序关联分析窗口
模糊划分	3 集合	低 / 中 / 高三级划分

将检测方法部署在 Dell PowerEdge R740 服务器上，配置双 Intel Xeon Gold 6248R 处理器和 4 块 NVIDIA T4 GPU。网络流量通过分光镜像方式接入检测系统，确保不影响业务正常运行。

4.3 应用效果分析

采用混淆矩阵对监测效果进行全面评估，结果如图 2 所示。

实际/预测	正常	侦查	入侵	横向移动	数据窃取
正常	99.2	0.3	0.2	0.2	0.1
侦查	1.8	97.5	0.4	0.3	0.0
入侵	0.7	1.2	96.8	1.1	0.2
横向移动	0.5	0.8	2.1	96.2	0.4
数据窃取	0.3	0.2	0.9	1.6	97.0

图 2 多阶段攻击检测混淆矩阵（%）

实验采用十折交叉验证，指标选取平均值。从图 2 多阶段攻击检测混淆矩阵中可以看出整体准确率达到 98.7%，这

意味着在所有的检测样本中，有 98.7% 的样本被正确分类，包括正常流量和各类攻击流量。这一高准确率表明本文提出的检测方法在整体上能够有效地对网络流量进行准确判断，具备较高的检测可靠性。误报率仅为 0.45%，即正常流量被错误检测为攻击流量的比例非常低。在实际网络环境中，误报会给网络管理人员带来不必要的干扰和工作量，甚至可能导致对正常业务的误处理。本文检测方法如此低的误报率，说明其能够较好地区分正常流量和攻击流量，减少了对正常业务的误判，有助于维持网络的稳定运行。漏报率为 1.3%，表明该方法能有效捕获绝大多数攻击流量。漏报的攻击流量可能会在网络中持续造成危害，因此较低的漏报率对于网络安全至关重要。本文检测方法在控制漏报率方面表现出色，能够及时发现绝大多数的攻击行为，最大程度降低攻击对网络造成的潜在风险。各攻击阶段的详细检测指标记录如表 4 所示。

表 4 分阶段检测性能指标

攻击阶段	精确率 /%	召回率 /%	F_1 值 /%
侦查	97.8	97.5	97.6
初始入侵	96.5	96.8	96.6
横向移动	96.1	96.2	96.1
数据窃取	97.3	97.0	97.1

同时，在检测过程中，本文方法共挖掘出 127 条强关联规则，部分典型规则如表 5 所示。

表 5 典型关联规则记录表

规则 ID	关联规则	支持度	置信度
R17	端口扫描→服务探测	0.23	0.89
R45	SQL 注入→异常 DB 连接	0.18	0.92
R82	凭证窃取→ SMB 异常访问	0.15	0.85
R113	权限提升→数据批量下载	0.12	0.91

从表 5 中可以看出，规则 R17 显示当检测到端口扫描行为后，置信度达 0.89。表示当检测到端口扫描行为后，在相应条件下出现服务探测的概率高达 89%。这一规则有效反映了攻击者在侦察阶段的行为模式。规则 R113 显示支持度为 0.12，置信度为 0.91，表示当攻击者成功提升权限后，有 91% 的概率会进行数据批量下载，这是目标达成阶段攻击者窃取数据的典型行为关联。这些典型关联规则的挖掘，为深入理解多阶段攻击的行为逻辑和内在联系提供了有力依据，比单一阶段检测能够更全面、准确地把握攻击的整体流程，可有效应对复杂网络高级威胁。

5 结语

本文针对计算机网络多阶段攻击检测能力受限问题，提出了一种融合关联规则与深度学习的协同检测框架。通过特征工程优化，利用改进的 NetFlowMeter 工具和 FS-VT 降维算法有效解决了原始流量数据的高维异构问题，为检测模型提供了高质量的输入数据。在规则挖掘方面，提出的自适应支持度策略与模糊集合划分方法显著提升了攻击模式发现效率，能够精准刻画攻击阶段间的行为关联性。基于多阶段攻击关联图模型（MS-AGM）和混合残差网络（HRN）的协同检测机制，结合深度时空特征与规则评分，实现了高准确率与低误报率的平衡。此外，分级响应机制（阻断、限速、日志）在实际部署中展现了良好的实时性与业务兼容性。然而，该方法对加密流量的处理仍有不足，未来将聚焦轻量化模型部署、动态规则库优化及加密流量分析，以增强在复杂网络环境中的适用性。

参考文献：

[1] 杨宏宇 , 张豪豪 , 成翔 . 基于多尺度注意力特征增强的异常流量检测方法 [J]. 通信学报 , 2024, 45(11): 88-105.

[2] 顾伟 , 行鸿彦 , 侯天浩 . 基于网络流量时空特征和自适应加权系数的异常流量检测方法 [J]. 电子与信息学报 , 2024, 46(6): 2647-2654.

[3] 陈向效 , 崔鑫 , 杜秦 , 等 . 基于机器学习的异常流量检测模型优化研究 [J]. 计算机科学 , 2024, 51(S1):982-986.

[4] 王文博 , 刘绚 , 张博 , 等 . 基于协议特征的电力工控网络流量异常行为检测方法 [J]. 电力系统自动化 , 2023, 47(2): 137-145.

[5] 王坤 , 付钰 , 段雪源 , 等 . 基于深度学习的 SDN 异常流量分布式检测方法 [J]. 通信学报 , 2024, 45(11):114-130.

[6] 李道全 , 李腾 , 李玉秀 . 基于机器学习的 SDN 异常流量检测架构 [J]. 计算机工程与设计 , 2023, 44(7):1928-1936.

[7] 钟昱 , 黄振南 , 谢惠超 , 等 . 一种基于半监督学习的网络异常流量检测方法 [J]. 广西大学学报（自然科学版）, 2024, 49(3): 563-574.

[8] 陈亮 , 李志华 . 面向物联网僵尸网络多阶段攻击的异常流量检测方法 [J]. 计算机科学 , 2024, 51(8):379-386.

[9] 冷依凌 , 邹细勇 . 基于 CNN-BiBASRU-AT 的网络异常流量检测模型 [J]. 微电子学与计算机 , 2024, 41(1):93-99.

[10] 陈万志 , 任鹏江 , 王天元 . 因素空间背景基的流量异常检测基点分类方法 [J]. 电子测量与仪器学报 , 2024, 38(6): 84-94.

【作者简介】

孙月玲（1984—），女，江苏盐城人，本科，讲师，研究方向：计算机网络技术。

（收稿日期：2025-07-09 修回日期：2026-01-14）