计算机科学
COMPUTER SCIENCE

物联网数据流威胁致效机理研究

孙瑞杰，李鹏，朱枫

引用本文

孙瑞杰，李鹏，朱枫. 物联网数据流威胁致效机理研究[J]. 计算机科学，2025，52(6)：397-404.

SUN Ruijie, LI Peng, ZHU Feng. Study on Efficacy Mechanism for IoT Data Flow Threats[J]. Computer Science, 2025, 52(6): 397-404.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240400133

物联网数据流威胁致效机理研究

孙瑞杰[1]　李　鹏[1,2,3]　朱　枫[1]
1 南京邮电大学计算机学院　南京 210023
2 国家高性能计算中心南京分中心　南京 210023
3 南京邮电大学网络安全和可信计算研究所　南京 210023
(2023040512@njupt.edu.cn)

摘　要　随着物联网设备的数量呈爆炸性增长,针对物联网设备的攻击手段也开始变得多样且隐蔽。基于机器学习的检测方法已经得到广泛的研究,并具有巨大的潜力。然而,这些模型被认为是黑匣子,很难解释其分类结果,因此无法说明物联网威胁特有的手段与模式。为了解决这个问题,文中基于ATT&CK框架,构建了技术-特征字典,将攻击技术进行了流量的特征化描述;并构建了威胁-技术数据库,将网络威胁分解到了攻击技术层面。文中设计了基于致效机理的威胁检测模型,构建了实时流量特征矩阵,归纳了流量受到的攻击技术,将技术序列代入威胁-技术数据库,得到可能受到的威胁及其概率。实验结果表明,所提模型对于数据集中的威胁检测率高达99.595%,与传统方法效果相当,并且可以根据实验环境需要调节误报率,为分析人员提供了可靠的攻击路径解释。

关键词:物联网数据流;威胁检测;致效机理;ATT&CK框架
中图分类号　TP393.08

Study on Efficacy Mechanism for IoT Data Flow Threats

SUN Ruijie[1], LI Peng[1,2,3] and ZHU Feng[1]
1 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
2 Nanjing Center of HPC China, Nanjing 210023, China
3 Institute of Network Security and Trusted Computing of NUPT, Nanjing 210023, China

Abstract　With the explosive growth in the number of IoT devices, the means of attacking IoT devices have also become diverse and covert. Machine learning-based detection methods have been actively researched and shown great potential. However, these models are considered black boxes, making it difficult to explain their classification results and thus unable to explain the specific means and patterns of IoT threats. To address this issue, this paper constructs a technology-feature dictionary based on ATT&CK framework, characterizing attack techniques with traffic features, and builds a threat-technology database, decomposing network threats into the level of attack techniques. This paper designs a threat detection model based on an efficacy mechanism, constructs a real-time traffic feature matrix, summarizes the attack techniques suffered by the traffic, and inputs the technical sequence into the threat-technology database to obtain the possible threats and their probabilities. Experimental results show that the proposed model achieves a threat detection rate as high as 99.595% in the dataset, which is compared to traditional methods. Moreover, it can adjust the false positive rate according to the experimental environment and provides reliable attack path explanations for analysts.

Keywords　IoT data flow, Threat detection, Efficacy mechanism, ATT&CK framework

随着物联网设备的数量呈爆炸性增长,物联网所带来的价值与便利开始真正体现,针对物联网设备的攻击手段也开始变得多样且隐蔽。攻击者们希望通过窃取隐私、损害可用性等方式,来得到自己的某些利益。在这一背景下,针对物联网威胁攻击的检测与防御手段开始变得更加重要。

传统的网络威胁检测方法,几乎都需要建立一个标准的训练模型,那便要牵扯到数据集的预处理与训练,并通过机器学习与深度学习的算法建立分类模型。传统的入侵检测系统建立过程中会存在两个问题:1)训练数据集是否在现代具有代表性,涵盖面是否足够广泛;2)机器学习、深度学习一类不具备可

到稿日期:2024-04-17　返修日期:2024-10-08
基金项目:江苏省六大人才高峰高层次人才项目(RJFW-111);江苏省研究生科研与实践创新计划项目(KYCX24_1227)
This work was supported by the Six Talent Peaks Project of Jiangsu Province(RJFW-111) and Postgraduate Research and Practice Innovation Program of Jiangsu Province(KYCX24_1227).
通信作者:李鹏(lipeng@njupt.edu.cn)

解释性的黑盒算法是否可靠，是否具有普遍性。

基于上述问题，本文提出了一种基于威胁致效机理的可解释型白盒威胁检测模型，该模型采用 ATT&CK 框架，根据威胁的生成路径研究攻击的流量特征，对威胁检测的涵盖面以及普适性进行改进。算法可根据使用环境的需求调节误报率，并增加了可溯源以及可解释性质。

1 相关工作

物联网是由 RFID、智能传感器、通信技术和互联网协议等技术的最新发展所推动的。目前，主流的威胁检测系统主要使用机器学习、深度学习以及统计学的方法。威胁检测系统旨在识别网络流量中可能体现恶意活动、网络攻击或系统故障的异常模式和行为[1]。现有的研究主要关注 3 类问题：改进性能、应用场景，以及技术方法。

关注性能的研究一般从模型结构入手，对一些已知的机器学习算法进行改进，以提升模型的泛化性与鲁棒性。Anitha 等[2]为了说明物联网网络的脆弱性，对卷积神经网络(CNN)、长短期记忆(LSTM)和 CNN 组合进行了研究和比较，他们使用 Modbus 网络流量数据检测物联网网络攻击，准确率超过 99%。Diro 等[3]深入回顾使用机器学习开发异常检测解决方案以保护物联网系统的现有工作，并指出基于区块链的异常检测系统可以协作学习有效的机器学习模型来检测异常。Deorankar 等[4]为了减少物联网设备中出现的网络攻击，通过引入雾层来维护云上可用数据的安全性，有助于控制各种网络攻击，此外还研究了雾层的工作以及不同的异常检测技术来防止网络攻击。由于机器学习、深度学习算法的普适性较差，因此针对场景的研究也被提出。Gu 等[5]针对 5G 电力失陷终端威胁，建立了突发信任评估模型，充分利用安装在电力终端的零信任代理持续性地收集终端行为因素，从中提取突发因子并量化反映到信任值，以快速发现和阻断具有突发异常行为的失陷终端。最后，关注技术方法的研究主要针对机器学习的学习方式，即无监督、半监督、自监督或是有监督的机器学习训练过程。Ding 等[6]针对 DDoS 攻击的防御提出了两种新的流量基数和归一化网络流量熵估计策略，这些策略仅使用 P4 支持的操作并保证较低相对的误差。Vugrin 等[7]详细介绍了一个数学模型，该模型描述了 NIDS 执行数据包检查及其对恶意软件 C2 流量的检测。Haji 等[8]在对机器学习方法和物联网安全在各种类型的潜在攻击中的重要性进行了全面的文献综述之后，从攻击检测和异常检测方面对各种机器学习算法进行了比较，并引入了可能的基于机器学习的物联网保护技术。Xiao 等[9]将数据包分为头段和有效负载段，然后将其输入到由具有注意力机制的循环神经网络组成的编码器中，通过另一个编码器学习整个数据包的高级表示。

深度学习需要更多的数据支撑来实现更高的准确性，然而物联网缺乏大型真实世界数据集，这是将 DL 模型纳入物联网的主要障碍[10]。深度学习的另一个限制是其模型专注于分类，而许多物联网应用(如电力负荷预测、温度预测)需要在其分析核心进行回归，很少有工作试图用回归功能来丰富深度神经网络。此时，急需设计一种适用性更强且具备可解释性的威胁检测算法，即通过威胁的形成路径以及导致的效果，来推断出流量的安全状态。

Xiong 等[11]提出了一种基于企业 ATT&CK 矩阵的企业安全威胁建模语言，该语言采用 Meta Attack 语言框架设计，主要用于描述系统资产、攻击步骤、防御措施和资产关联，该语言中的攻击步骤代表了 MITRE 所列出和描述的对手技术。Al-Shaer 等[12]介绍了针对 ATT&CK 报告的 APT 和软件攻击数据进行的统计机器学习分析，以推断代表显著相关性的技术聚类，这些聚类可用于技术预测，并使用分层聚类来推断攻击技术关联，提供了具有统计显著性和可解释性的技术相关性。Georgiadou 等[13]利用 MITRE ATT&CK 框架，介绍了一套全面的组织和个人文化因素与映射到特定对手行为和模式的安全漏洞关联起来的研究结果。Kwon 等[14]提出了一种名为“网络威胁字典”的工具，通过将 MITRE ATT&CK 矩阵映射到 NIST 网络安全框架，为威胁提供了方法和实用解决方案。通过为网络安全从业者提供即时解决方案，网络威胁字典能够有效应对网络攻击。Haque 等[15]捕获攻击图的节点描述，并应用 TF-IDF 算法将攻击技术与可用的节点描述进行映射，生成余弦相似性，以确定对手攻击网络的方法。

然而，普适性的流量异常检测方法尚未被提出，对照于 ATT&CK 框架中的攻击流程，一个标准的威胁流量刻画方案对于威胁致效机理的研究是必要的。针对这些问题，本文基于 ATT&CK 框架，提出了威胁-技术数据库来表示威胁的分解，构建了技术-流量字典来实现用流量特征描述技术，并建立了基于致效机理的威胁检测模型，在适用性以及可解释性方面做出了优化。

2 基于致效机理的物联网威胁检测模型设计

对于网络状况的异常与否，从代码层面进行研究是十分困难的，从威胁的形成原理以及影响效果角度对其进行观测是一种直观有效的方法。其中，可通过网络流量的特征来映射出一些威胁的致效机理。例如，当网络受到拒绝服务攻击时，系统部分功能会瘫痪，反映出的流量特点即为输入流量包的数量会大幅增加；当受到一些协议攻击时，数据包协议特征会出现异常，当受到中间人攻击或者扫描、嗅探等威胁时，源 IP 应当不存在于访问记录表中。因此，对于网络威胁检测的研究可以从致效机理角度出发，从流量特征的角度体现致效特点，根据相应特征的特点来判断网络的安全状态。

2.1 基于 ATT&CK 技术的威胁分解

ATT&CK 框架认为，一种威胁的实现并非一步就能直接达到其攻击目的，而是通过一系列的步骤，即“技术”，从而完成其威胁目的。因此，想要研究一类威胁，一个较好的解决方法即为将威胁进行分解，从“技术”层面对威胁进行更细致化的研究。

本文主要从流量角度探测流量的安全状态，因此需要先提取出可从流量角度上检测到的 ATT&CK 技术。ATT&CK 技术框架中包含了关于某一技术的程序用例、缓解措施以及探测方法描述。缓解措施表示安全概念和技术

类别，可用于阻止技术或子技术成功执行。探测方法根据数据源出发，表示对该威胁技术的发现预防措施。

缓解措施包括应用程序隔离和沙箱、数据备份、执行预防和网络分段等措施，其中网络流量过滤（Filter Network Traffic）即使用网络设备过滤入口或出口流量，并执行基于协议的过滤，在端点上配置软件以过滤网络流量，即从流量特征的角度阻断部分攻击。

探测方法包含固件、图像、运营数据库等数据源，其中网络流量（Network Traffic）表示通过网络传输的数据（如Web、DNS、邮件、文件等），这些数据要么被汇总，要么以可分析格式被捕获为原始数据，包含网络连接创建、网络流量内容、网络流量流3种检测角度。

因此，对于流量特征的分析，需要用到可通过流量角度探测的技术，即在“缓解措施”阶段包含网络流量过滤，或在“探测方法”阶段包括网络流量的技术。经过汇总，具体的技术名称如图1所示。

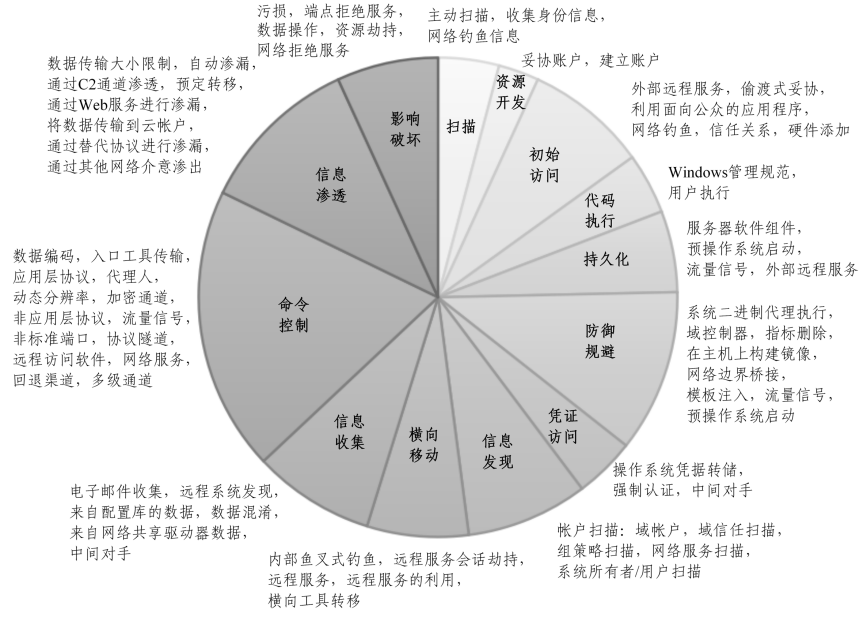


图1 可通过流量角度检测的ATT&CK技术统计图

Fig. 1 Statistics of ATT&CK detectable techniques from traffic perspective

2.2 ATT&CK技术字典构建

在ATT&CK框架中，每一种技术的描述都是现实世界的映射，因此对于每一种技术，ATT&CK都会记录其程序示例、缓解措施、检测方法以及相关参考文献，经过检测方法的描述，可以得到对于一种技术的检测中所需重点关注的流量特征类型。在展开相关实验之后，即可得到在该种技术的供给情况下，流量特征的具体变化情况。基于此想法，本文提出构建“技术-特征字典”，即建立一个矩阵模型，其中每一行代表一种ATT&CK技术的流量特征，实现用流量特征来表示特定的ATT&CK技术。

观察现阶段各种数据集的特征取值方式，可以发现特征共可以分为三大类，即非数值类型特征、数值类型但与数值大小无关特征，以及数值类型且与数值大小有关特征。其具体含义与分类方法如表1所列。

表1 流量特征类别

Table 1 Categories of traffic features

类别	描述	特征举例
A	非数值类型特征	flgs，proto，state
B	数值类型但与数值大小无关特征	saddr，daddr，sport，dport
C	数值类型且与数值大小有关特征	Dintpkt，Dpkts，tcprtt，ct_dst_ltm，

对于处于期望状态下的A特征，可以标记其为1，表示该特征项正常；对于非期望状态下的特征，可以标记其为0，表示该特征项异常。对于B特征，构建一个列表记录常用值，若特征项的值处于记录列表中，则标记其为1，若特征项的值不处于记录列表中，则标记其为0。对于占大多数的C特征，对常规流量数据进行截取分析，统计其正常值范围，超过范围的数值定义为异常数值，正常则记录为1，异常则记录为异常系数，即越过正常取值范围的倍数，超过范围最大值100倍为无穷，低于范围最低值0.01倍为0。对于流量特征$t_{i,j}$的取值类型及取值范围如式(1)所示：

$$t_{i,j}=\begin{cases}0, & A,B\text{类特征异常},C\text{类特征远低于正常值}\\1, & A,B,C\text{类特征正常}\\k, & C\text{类异常特征的异常系数}\end{cases}\tag{1}$$

对于2.1节所统计的65种不同技术，以及现阶段各类数据集中所提供的3类流量特征，可以构建3个矩阵来表示3类特征，以显示流量之间的关系，即技术-特征字典。其表现形式如式(2)所示：

$$\boldsymbol{T}_i=\begin{pmatrix}t_{1,i_1} & t_{1,i_2} & \cdots & t_{1,i_a}\\ t_{2,i_1} & t_{2,i_2} & \cdots & t_{2,i_a}\\ \vdots & \vdots & \cdots & \vdots\\ t_{65,i_1} & t_{65,i_2} & \cdots & t_{65,i_a}\end{pmatrix},i\in\{A,B,C\}\tag{2}$$

其中，a表示3类特征各自对应的特征数量，矩阵$\boldsymbol{T}_A$，$\boldsymbol{T}_B$，$\boldsymbol{T}_C$同为65行，表示的是65种可通过流量角度检测到的技术。

对于 $t_{i,j}$ 的具体值，需要根据流量特征的具体类型来决定，具体如式(1)所示。经过以上流量类别的讨论，可对“技术-特征字典”进行优化，即将特征矩阵扩充为 3 个高度相等的矩阵，3 个矩阵的每一行共同来刻画一种 ATT&CK 技术。

构建流量特征字典时，主要是用特征来对 ATT&CK 技术进行描述，该类技术被现有的公开流量特征数据集所包含并描述，可直接对数据集中的特征进行归纳统计。Bagui 等在 2023 年正式发布了 UWF-ZeekData22 数据集，该数据集使用 Zeek 创建，并使用 MITRE ATT&CK 框架标记。该数据集可用于检测导致攻击的对手行为，开发意图进行攻击的用户或用户组的配置文件以及识别攻击流量和攻击。

对于数据集有利用价值的信息而言，需要提取出其特征数据的特点，现有的机器学习算法一般只可提供分类结果，很难直接提供分类依据，因此对其特征可采取统计学的方法进行研究。均值、取值范围这两个统计量对于本文研究较为重要，通过均值与取值范围构建出一个常规取值范围量，即可完成有数据集技术特征字典的构建。

2.3 威胁-技术数据库构建机理

ATT&CK 框架作为一个综合性的网络威胁攻击库，包含了众多有关威胁攻击的理论知识及探测方法。ATT&CK 框架认为，所有威胁的实现必然会经历一系列步骤，即侦查、资源开发、初始访问、代码执行、持久化、提权、防御规避、凭证访问、信息发现、横向移动、信息收集、命令控制、信息渗透、影响破坏 14 个战术。网络攻击的产生往往不是在一瞬间完成的，而是会经历一个完整的攻击过程，每一过程都需要通过特定的手段去完成。基于此，本文构建了“威胁-技术数据库”，即通过 ATT&CK 中的技术去刻画某一次网络威胁，其体现形式是一个矩阵，行数即为数据库涵盖的威胁数量，每一行包括 65 个元素，分别代表 65 种 ATT&CK 技术，代表一类威胁的攻击流程与方式。

其中，每一种战术都有其对应的实现技术，即一种威胁若经历该战术阶段，则必然会使用到该战术所覆盖的其中一种技术手段。在考虑网络流量探测的情况下，相关的技术如图 1 所示。既然一个威胁的实现必然会使用到一系列的技术手段，那么可以构建一个技术矩阵，用其来表示一种威胁的分解，如式(3)所示：

$$\boldsymbol{A}'=\begin{pmatrix} s_{1,1} & s_{1,2} & s_{1,3} \\ s_{2,1} & & s_{2,2} \\ s_{3,1} & \cdots & s_{3,6} \\ s_{4,1} & & s_{4,2} \\ s_{5,1} & \cdots & s_{5,4} \\ s_{6,1} & \cdots & s_{6,8} \\ s_{7,1} & s_{7,2} & s_{7,3} \\ s_{8,1} & \cdots & s_{8,5} \\ s_{9,1} & \cdots & s_{9,5} \\ s_{10,1} & \cdots & s_{10,6} \\ s_{11,1} & \cdots & s_{11,14} \\ s_{12,1} & \cdots & s_{12,8} \\ s_{13,1} & \cdots & s_{13,5} \end{pmatrix} \tag{3}$$

在 $\boldsymbol{A}'$ 中，共存在 13 行。由于提权战术所涉及的技术无法通过流量角度检测，因此删去这一种战术阶段，剩下的 13 种战术依次分别对应矩阵的每一行，每行中元素的个数即为该类战术中可通过流量方面检测到的技术的数量。我们根据现有网络威胁流量数据集，选取威胁的标签记录点前 50 条流量数据，与技术-特征字典中的所有数据进行匹配，计算其 cos 距离相似度，计算方法如式(4)所示：

$$\begin{aligned} b_{Aji}&=1-\frac{\sum_{k=1}^{a} r_{j,A_k}\oplus t_{i,A_k}}{a} \\ b_{Bji}&=1-\frac{\sum_{k=1}^{b} r_{j,B_k}\oplus t_{i,B_k}}{b} \\ b_{Cji}&=1-\frac{\sum_{k=1}^{a} r_{j,A_k}\oplus t_{i,A_k}}{a} \\ b_{ji}&=we_A\cdot b_{Aji}+we_B\cdot b_{Bji}+we_C\cdot b_{Cji} \\ we_A&+we_B+we_C=1 \end{aligned} \tag{4}$$

A 和 B 类特征类型皆为布尔型数据，因此与流量-特征字典中的相应特征值进行异或取非运算即可；C 类特征为浮点型数据，因此对其先进行归一化，再采取计算 cos 距离相似度的方法进行相似度研究。每个时间段中选取最大的 b 值作为本时间段的技术相似度 w，对整个时间段中的流量进行比对后，选取每一技术的最大 w 值作为威胁 A 的技术表征，将单一威胁构建为行列式 A，如式(5)所示：

$$\begin{aligned} w_i&=\max_j b_{ji} \\ A&=(\max(w_1)\ \max(w_2)\ \cdots\ \max(w_{65})) \end{aligned} \tag{5}$$

完整的行列式 A 的计算方法如算法 1 所示。

算法 1 威胁-技术行列式 A 构建算法

输入：traffic_data(流量数据集)，T(技术-特征字典)

输出：A(威胁-技术行列式)

```
1. 截取威胁标签记录点前 50 条流量数据
   selected_ data=traffic_data [label-50:label]
2. 定义布尔型数据相似度计算函数
   def bool_similarity(r,t):
3.     sam=sum(r[k] ^ t[k] for k in range(len(r)))
4.     return(1-sam)/ len(r)
5. 定义浮点型数据相似度计算函数
   def float_similarity(r,t):
6.         sam=sqrt(sum((r[k] - t[k])²for k in range(len(r))))
7.     return(1-sam)/ len(r)
8. b_Aji=bool_similarity(r['A'],t['A'])
9. b_Bji=bool_similarity(r['B'],t['B'])
10. b_Cji=float_similarity(r['C'],t['C'])
11. 计算加权相似度 b_ji
    b_ji=w_A * b_Aji+ w_B * b_Bji+ w_C * b_Cji
12. 对每一种技术计算相似度
    tech_features=T[tech]
13. if b_ji> max_b:max_b=b_ji
14.     w[tech]=max_b
15. 更新威胁-技术行列式 A
    for i in range(num_tech_features)]:
16.     A_i=[max(w[tech][i] for tech in range(65))。
```

若存在 n 种已知威胁，则对所有威胁进行上述分解，最后

可得到一个威胁-技术数据库，如式(6)所示：

$$\boldsymbol{W}=\begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,65} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,65} \\ \vdots & \vdots & \cdots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,65} \end{pmatrix} \tag{6}$$

在 $\boldsymbol{W}$ 中，共存在 n 行、65 列，分别表示 n 种已知威胁以及 65 项 ATT&CK 模型中可通过流量角度检测到的技术。矩阵中的每一个变量 $w_{i,j}$ 是一个离散型变量，$w_{i,j}$ 的数值表示第 i 种威胁会使用到第 j 种技术的概率。以 DoS 攻击与 Backdoors 攻击为例，具体的分解结果如表 2 所列。

表 2 威胁的攻击技术分解

Table 2 Decomposition of attack techniques for DoS threats

威胁名	经历战术阶段	对应攻击技术
拒绝服务攻击(DoS)	初始入侵	外部远程服务
	持久化	外部远程服务
	信息发现	网络服务扫描
	信息收集	电子邮件收集
	影响破坏	网络拒绝服务
后门攻击(Backdoors)	初始入侵	面向公众应用程序
	持久化	流量信号
	防御规避	流量信号
	横向移动	远程服务
	命令控制	远程访问软件
	信息渗透	自动渗透
	影响破坏	资源劫持

2.4 基于致效机理的威胁检测方法

若想实现一个网络威胁检测系统，还需要将威胁真正的联系到流量特征的角度上。基于 ATT&CK 框架，在 2.1 节中完成了威胁的技术分解，在 2.2 节中完成了流量检测角度相关技术的流量字典构建，对两者的工作进行结合，即可得到最终的威胁-流量数据库。该数据库从致效机理的角度出发，比较完善地描述了在现实世界中各类威胁手段所产生的流量特征方面的影响。

ATT&CK 模型认为，一次攻击的实行将会依照模型的战术步骤逐渐展开，每经历一战术步骤，必然会使用到该战术层面下的一种技术。对一段时间的流量包进行特征提取，每一个数据包可被分解为一系列的特征集合，将特征集合按照式(1)分为 3 个特征类别后，可得到该段时间内的 3 个流量特征矩阵，如式(7)所示：

$$\boldsymbol{R}_i=\begin{pmatrix} r_{1,i_1} & r_{1,i_2} & \cdots & r_{1,i_a} \\ r_{2,i_1} & r_{2,i_2} & \cdots & r_{2,i_a} \\ \vdots & \vdots & \cdots & \vdots \\ r_{n,i_1} & r_{n,i_2} & \cdots & r_{n,i_a} \end{pmatrix}, i\in\{A,B,C\} \tag{7}$$

在流量特征矩阵 $\boldsymbol{R}$ 中，共有 n 行，n 即该段时间内流量数据包的数量，其中 a 表示 3 类特征各自对应的特征数量。将 $\boldsymbol{R}$ 中每段流量 R_i 与技术-特征字典 T 的每一行进行比对，相似度最高的即为这段流量理论上所受的攻击技术情况。

为了判断研究的流量具体属于哪个 ATT&CK 技术类别，需要对其与技术-特征字典中的内容进行相似度比对，以研究其威胁情况，具体比对方法如式(4)所示。完成整个时间段中的流量分析比对后，可以得到这段流量整体所受到的攻击技术种类，将其总结为一个列向量，如式(8)所示：

$$b_i=\max_j b_{ji}$$

$$\boldsymbol{B}=(b_1\ \ b_2\ \ b_3\ \ \cdots\ \ b_{65})^{\mathrm{T}} \tag{8}$$

列向量 $\boldsymbol{B}$ 含有 65 个元素，每个元素 b_i 是一个取值范围为 0 到 1 的浮点数，代表该段流量收到第 i 种 ATT&CK 技术威胁的可能性。b_i 越靠近 0 时，表示该段流量受到第 i 种技术攻击的可能性越低；b_i 越靠近 1 时，表示该段流量受到第 i 种技术攻击的可能性越高。最后，还需要探索该段流量所属的具体威胁类型，将威胁-技术字典 $\boldsymbol{W}$ 的每一行与 $\boldsymbol{B}$ 进行下述的概率和运算，最后得到一个判别向量，如式(9)所示。完整的判别向量 $\boldsymbol{J}$ 的计算方法如算法 2 所示。

$$\begin{aligned}\boldsymbol{J}&=\boldsymbol{W}\cdot\boldsymbol{B}\\&=\begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,65} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,65} \\ \vdots & \vdots & \cdots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,65} \end{pmatrix}\cdot\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{65} \end{pmatrix}\\&=\begin{pmatrix} \sum_{i=1}^{65} w_{1,i}*b_i \\ \sum_{i=1}^{65} w_{2,i}*b_i \\ \vdots \\ \sum_{i=1}^{65} w_{n,i}*b_i \end{pmatrix}=(j_1\ \ j_2\ \ j_3\ \ \cdots\ \ j_n)^{\mathrm{T}}\end{aligned} \tag{9}$$

$j_i\to 0$，表示该段流量处于第 i 种威胁的概率越小；$j_i\to 65$，表示该段流量处于第 i 种威胁的概率越大。

算法 2 威胁判别向量 $\boldsymbol{J}$ 的计算算法

输入：W(威胁-技术字典)，Sim_matrix(相似度矩阵)

输出：J(威胁判别向量)

1. 计算每个技术的最大相似度

   for j in range(num_techs)：

2. B[j]=max(similarity_matrix[i][j] for i in range(num_rows))

3. 计算威胁判别向量 J

   for i in range(num_threats)：

4. J[i]=sum(W[i][k] * B[k] for k in range(num_techs))

在该运算方法中，对于第 i 类威胁而言，若其不涉及第 j 种技术，则希望 b_j 尽可能小，若其设计第 j 种技术，则希望 b_j 尽可能大。因此在 $w_{i,j}$ 取 0 时，b_j 的值取为 $(1-b_i)$；在 $w_{i,j}$ 取 1 时，b_j 的值不做改变。最后对 b_j 进行求和运算，即可得到该段流量受到第 i 种威胁的概率。

判别向量 $\boldsymbol{J}$ 共含有 n 个元素，j_i 表示与第 i 种威胁的相似度。在认为该段流量为异常流量的基础上，j_i 的值越接近 0 代表该段流量处于第 i 种威胁的概率越大，j_i 的值越接近 65 代表该段流量处于第 i 种威胁的概率越小。

在窗口大小的取值方面，需要根据具体情况进行分析。首先需要设定一个合适的标准窗口大小值，当一段流量被检测出受到技术攻击时，窗口开始；当一段流量受到的检测战术为“影响破坏”阶段时，该窗口立即结束；若在标准窗口大小达到后，仍没有检测出“影响破坏战术”，则此窗口结束。总结出的最可能的威胁攻击类型标注在最后一段被检测出攻击技术的流量段之处。

综上所述，威胁-流量数据库构建的原理即是从威胁的致效机理出发，基于 ATT&CK 框架，对威胁进行风险评估，找

到该威胁所需经历的战术阶段，再进一步找到其所采用的技术手段，将其映射入 $\boldsymbol{W}'$ 矩阵中，然后根据调研、实验得到的技术-流量字典，将一类已知的威胁真正地完全分解成具体的流量特征，并将其加入到威胁-流量数据库。在完成所有需要研究的威胁评估后，威胁-特征数据库 WDB 即构建完成。将经过数据处理得到的流量特征信息代入数据库后，即可得到反馈的威胁评估种类以及其实现概率。

3 基于致效机理的物联网威胁检测模型测试

3.1 测试环境与测试内容

本文使用 MacOS 作为实验开展的运行操作系统，使用 Python 编程语言进行代码编写，使用 PyQT5 作为界面工具。整个测试环境参数如表 3 所列。本模型设计了基于致效机理的物联网威胁检测方法，测试环节主要针对测试结果的说明，输出测试进行。本节将对用于测试的数据集进行介绍，程序主要包含 3 个步骤，即基础搭建、技术分析、威胁归纳。具体如下。

首先，需要构建 3 个列表，即包含威胁名的威胁列表、包含技术名的技术列表，以及包含特征名的特征列表，用于在输出时找到对应威胁或技术的名称。接着，创建两个二维列表，分别代表标准的威胁-技术数据库，以及技术-流量字典。

然后，对数据集中的每一条数据进行分析比对，先对应到技术-流量字典，依次与每一种技术进行相似性比对，若相似度大于 60%，则认为该段流量受到技术的攻击，并在结果中输出相似度最高的技术名及对应的相似度。

在 ATT&CK 框架中，“影响破坏”战术位于最后阶段，且本次实验所涉及的威胁最终会起到“影响破坏”的攻击目的。因此对于窗口大小的选取，当一类攻击技术被检测出时，窗口开始，当对于“影响破坏”战术阶段中对应技术的检测相似度达到 80%以上时，窗口结束。对于窗口中的流量，在窗口末端进行威胁统计，对该窗口内的攻击技术以及相似度输送到威胁-技术数据库并按照式(9)进行比对，得出最可能收到的威胁类别以及对应的概率。

表 3 测试环境参数

Table 3 Test environment parameters

参数		环境
主机环境	操作系统	MacOS Monterey 12.3
	处理器	2.8 GHz 四核 Intel Core i7
	内存	16 GB 2133 MHz LPDDR3
编程环境	编程语言	Python 3.10
	界面工具	PyQT5
	编程软件	PyCharm CE

程序最后以 xlsx 文件形式进行输出，同时还对威胁检测算法的准确率、精确度、漏报率、误报率等性能指标进行评估，并与经典的流量分类算法的性能进行了对比。

3.2 数据集

UNSW-NB15 数据集[16]是广泛用于评估异常流量检测系统的基准数据集。它由澳大利亚新南威尔士大学的网络安全实验室开发。该数据集通过使用 UNSW-Network Telescope 在受控环境中捕获网络流量来收集。该望远镜由大量未使用的 IP 地址组成，仅接收非请求的流量，为研究恶意活动提供了独特的来源。该数据集涵盖了各种网络流量种类，包括正常流量、网络攻击和异常流量。它包括 9 种不同类型的攻击，如 DoS、探测、R2L 和 U2R，模拟真实场景。

UNSW-NB15 数据集已被广泛用于研究人员开发和评估异常流量检测系统，以及比较不同的算法，并根据准确性、精确度、召回率和其他评估指标来评估算法性能。

3.3 输出结果演示

对于检测生成的文件，其中序号列表示流量序号，原标签列表示原数据集对其打上的攻击类别标签，所受攻击技术列表示算法检测该段流量受到的攻击技术情况，相似度表示该段流量与技术特征的相似度，威胁名表示算法归纳出的窗口内流量受到的威胁种类，可能性表示受到该类威胁的概率。

实验主要针对 DoS 攻击以及 Backdoors 攻击进行，因此对两种威胁分别采取一条分析结果来具体说明。首先对 DoS 攻击的检测结果进行说明，如表 4 所列，第 2099 条流量与“网络拒绝服务”技术有 77.7%的相似度。对于序号为 2080 到 2099 的流量，DoS 攻击的分解技术“外部远程服务”“电子邮件收集”“网络服务扫描”“网络拒绝服务”在这个窗口能被全部检测。在窗口结束的位置，定位此段流量受到了 DoS 攻击，且概率为 86.9%。

表 4 DoS 攻击检测结果说明

Table 4 Explanation of DoS attack detection results

序号	原标签	所受攻击技术	相似度/%	威胁名	可能性/%
2075		流量信号	92.8		
2078		远程访问软件	66.7		
2080		电子邮件收集	72.8		
2081		外部远程服务	92.4		
2082		电子邮件收集	60.0		
2088		外部远程服务	78.3		
2092		资源劫持	94.1	DoS	79.8
2094		网络服务扫描	87.8		
2097		自动渗透	63.8		
2098		外部远程服务	81.8		
2099	DoS	网络拒绝服务	77.7	DoS	86.9

对 Backdoors 攻击的检测结果如表 5 所列，第 88313 条流量与“资源劫持”技术有 85.9%的相似度。对于序号为 88294 到 88313 的流量，Backdoors 攻击的分解技术“流量信号”“面向公众的应用程序”“远程访问软件”“远程服务”“自动渗透”在这个窗口能被全部检测。在窗口结束的位置，定位此段流量受到了 Backdoors 攻击，且概率为 80.0%。

表 5 Backdoors 攻击检测结果说明

Table 5 Explanation of Backdoor attack detection results

序号	原标签	所受攻击技术	相似度/%	威胁名	可能性/%
88290		流量信号	63.2		
88291		流量信号	67.8		
88292		面向公众的应用程序	61.1		
88293		资源劫持	80.6		
88294		远程访问软件	89.8		
88296		网络拒绝服务	94.2		
88297		外部远程服务	60.1		
88300		远程访问软件	88.5		
88306		远程访问软件	78.6		
88307		远程访问软件	60.3		
88308		自动渗透	87.8		
88309		自动渗透	61.5		
88310		远程服务	66.8		
88311		流量信号	81.5		
88313	Backdoors	资源劫持	85.9	Backdoors	80.0

4 算法性能分析

4.1 模型性能分析

异常检测算法的性能指标一般有：准确率、精确率、召回率、误报率、漏报率、特异度。将异常流量预测正确的记为TP，异常流量预测错误的记为FN，正常流量预测为异常样本的记为FP，正常流量预测为正常样本的记为TN。准确率Acc、精确率Pre、误报率Fpr、召回率Rec、漏报率Mis、特异度Tnr的计算式如式(10)－式(15)所示：

$$Acc=\frac{FN+TN}{FN+TN+TP+FP} \tag{10}$$

$$Pre=\frac{TP}{TP+FP} \tag{11}$$

$$Fpr=\frac{FP}{TN+FP} \tag{12}$$

$$Rec=\frac{TP}{TP+FN} \tag{13}$$

$$Mis=\frac{FN}{TP+FN} \tag{14}$$

$$Tnr=\frac{TN}{FP+TN} \tag{15}$$

在模型中，性能指标的数值主要取决于两点：字典的准确性与可接受异常概率的阈值。字典的准确性问题在于，威胁的技术分解是否足够合理全面以及技术的流量化表示是否明确清晰，这方面的影响在短期内难以修改到更高的技术水平。因此，对于模型性能的影响点主要在于可接受异常的概率阈值。在可接受的异常概率最低值为70%时，漏报率为0。因此，修改阈值范围为70%～90%，观察模型的性能变化范围。在本威胁检测模型中，误报率是较为重要的一个性能指标，代表着对威胁的识别程度，因此绘制漏报率与其他指标的折线图，来展示当可接受的异常概率最低值发生变化时，模型的性能情况。

由式(13)和式(14)可知，漏报率与召回率之和为1，因此不再展示它们之间的关系图像。本文模型的检测准确度较高，当漏报率为0时，检测准确率仍能达到98.4%以上。漏报率与准确率的具体关系图如图2所示。

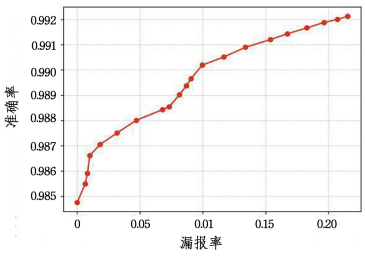


图2 漏报率与准确率的数值关系折线图

Fig. 2 Line chart of numerical relationship between false negative rate and accuracy rate

由式(12)和式(14)可知，漏报率与误报率之和为1，因此只展示漏报率与误报率之间的关系。当漏报率为0时，模型的误报率也保持在0.6%以下。漏报率与误报率的具体关系图如图3所示。

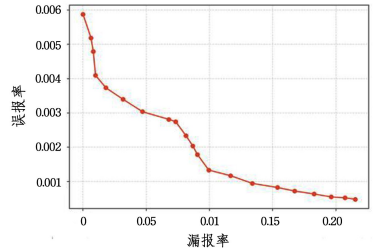


图3 漏报率与误报率的数值关系折线图

Fig. 3 Line chart of numerical relationship between false negative rate and false positive rate

模型的精确率与漏报率是成正比的，即在得到高精确率的同时也会带来较高的漏报率。在本实验中，由于异常流量相对于正常流量数量过少，因此精确度与漏报率是最重要且最能反映出模型性能的性能指标。漏报率与精确度的具体关系图如图4所示。

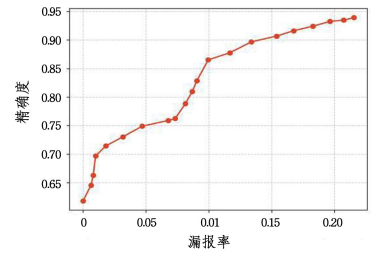


图4 漏报率与精确度的数值关系折线图

Fig. 4 Line chart of numerical relationship between false negative rate and accuracy

4.2 算法对比分析

考虑到计算资源与设备的影响，本文选取机器学习算法(随机森林、支持向量机(SVM))进行对比实验。对于本文基于致效机理的威胁监测模型，我们选取可接受的异常概率最低值为75%，此时本文模型的漏报率与精确度相对平衡，统计两种算法的检测结果指标。，具体数据如表6所列。

表6 算法检测结果对比

Table 6 Comparison of algorithm detection results

(条)

算法名	TP	FN	FP	TN
基于SVM的威胁检测模型	1070	36	592	115695
基于致效机理的威胁检测模型	1038	68	408	115879
基于随机森林的威胁检测模型	1013	93	1861	114426

对两种算法的准确率、精确度、召回率、误报率、漏报率、特异度6个性能指标进行计算，结果如表7所列。

表7 算法性能指标对比

Table 7 Comparison of algorithm performance metrics

(%)

算法名	准确率	精确度	召回率	漏报率	误报率	特异度
基于SVM的威胁检测模型	99.465	64.380	96.745	3.255	0.509	99.491
基于致效机理的威胁检测模型	99.595	71.784	93.852	6.148	0.351	99.649
基于随机森林的威胁检测模型	98.336	35.247	91.591	8.409	1.600	98.000

由表 7 可知，相比随机森林，本文算法在各个性能指标角度皆有优势。相比 SVM，本文算法的精确度与误报率具有优势，但是召回率与漏报率具有劣势。本文模型是专门针对威胁检测设计的，针对威胁检测系统具有更好的适用性，且可根据需要，调动可接受的异常概率最低值。针对安全需求高的系统，可对其进行调高；对于性能较差的系统，可对其进行调低。相比于机器学习的参数调整，本文模型可输出一次威胁行为的完整流程与攻击方法，其分类结果更具有可解释性。

结束语 本文主要研究基于致效机理的网络威胁检测，通过分析和评价前人的网络威胁检测模型，提出自己的模型并进行测试。具体工作如下：

1)针对前人提出的基于机器学习、深度学习的威胁检测模型进行了总结和分析，通过总结这些模型的特点和优劣，提出了可解释性算法模型的必要性，说明了基于致效机理的威胁检测模型的优势。

2)调研了 ATT&CK 框架模型的结构以及用途，分析总结了可通过流量角度实现检测的攻击技术。根据文献说明以及流量信息，构建了威胁-技术字典以及技术-特征字典，实现了威胁的技术化分解，以及攻击技术的流量化体现。

3)根据采集的流量或现有数据集，给出了基于致效机理的威胁评价模型。模型中可自主设置接受的异常概率阈值。

4)最后，在实现上述基于致效机理的威胁检测模型的软件程序后，本文对上述检测流程进行了测试，说明了测试结果的可解释性与溯源过程，并与传统的机器学习算法进行了检测效果的对比分析，进一步验证了本文算法的优秀性能。

参 考 文 献

[1] SHANG Y, LI P, ZHU F, et al. Overview of IoT traffic attack detection technology based on fuzzy logic[J]. Computer Science, 2024, 51(3): 3-13.

[2] ANITHA V, KUMAR C G N, KUCHIPUDI R, et al. Cybersecurity in Internet of Things Networks using Deep Learning Models[C]// 2023 International Conference on Sustainable Computing and Data Communication Systems(ICSCDS). IEEE, 2023: 1090-1095.

[3] DIRO A, CHILAMKURTI N, NGUYEN V D, et al. A comprehensive study of anomaly detection schemes in IoT networks using machine learning algorithms[J]. Sensors, 2021, 21(24): 8320.

[4] DEORANKAR A V, THAKARE S S. Survey on anomaly detection of(iot)-internet of things cyberattacks using machine learning[C]// 2020 Fourth International Conference on Computing Methodologies and Communication(ICCMC). IEEE, 2020: 115-117.

[5] GU Z, WANG Z, GUO J, et al. 5G power failure terminal threat detection based on atomized zero-trust component[J]. Computer Engineering, 2023, 49(2): 161-168.

[6] DING D, SAVI M, SIRACUSA D. Tracking normalized network traffic entropy to detect DDoS attacks in P4[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 19(6): 4019-4031.

[7] VUGRIN E D, HANSON S, CRUZ J, et al. Experimental Validation of a Command and Control Traffic Detection Model[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 21(3): 1084-1097.

[8] HAJI S H, AMEEN S Y. Attack and anomaly detection in IoT networks using machine learning techniques: A review[J]. Asian Journal of Research in Computer Science, 2021, 9(2): 30-46.

[9] XIAO X, XIAO W, LI R, et al. EBSNN: Extended byte segment neural network for network traffic classification[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 19(5): 3521-3538.

[10] QIU X, ZHANG L, REN Y, et al. Ensemble deep learning for regression and time series forecasting[C]// 2014 IEEE Symposium on Computational Intelligence in Ensemble learning (CIEL). IEEE, 2014: 1-6.

[11] XIONG W, LEGRAND E, ÅBERG O, et al. Cyber security threat modeling based on the MITRE Enterprise ATT&CK Matrix [J]. Software and Systems Modeling, 2022, 21(1): 157-177.

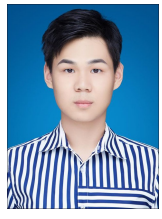
[12] AL-SHAER R, SPRING J M, CHRISTOU E. Learning the associations of mitre att & ck adversarial techniques[C]// 2020 IEEE Conference on Communications and Network Security (CNS). IEEE, 2020: 1-9.

[13] GEORGIADOU A, MOUZAKITIS S, ASKOUNIS D. Assessing mitre att&ck risk using a cyber-security culture framework[J]. Sensors, 2021, 21(9): 3267.

[14] KWON R, ASHLEY T, CASTLEBERRY J, et al. Cyber threat dictionary using mitre att&ck matrix and nist cybersecurity framework mapping[C]// 2020 Resilience Week(RWS). IEEE, 2020: 106-112.

[15] HAQUE M A, SHETTY S, KAMHOUA C A, et al. Adversarial Technique Validation & Defense Selection Using Attack Graph & ATT&CK Matrix[C]// 2023 International Conference on Computing, Networking and Communications(ICNC). IEEE, 2023: 181-187.

[16] NOUR M. The UNSW-NB15 Dataset [EB/OL]. https://paperswithcode.com/dataset/unsw-nb15.



SUN Ruijie, born in 2000, postgraguate. His main research interests include network traffic security and watermark of large language models.

LI Peng, born in 1979, Ph.D, professor, Ph.D supervisor, is a member of CCF (No. 48573M). His main research interests include computer communication networks, cloud computing and information security.

(责任编辑：喻藜)