

DOI: 10.3969/j.issn.1671-0673. 2026. 02. 012

SAFE-Former: 一种针对 SSH 协议的 中间人攻击检测方法

刘 骥, 林 伟, 靳 迪, 李 玘, 陈 迪

(信息工程大学, 河南 郑州 450001)

摘要: 针对 SSH(Secure Shell)协议中间人攻击检测中传统方法泛化能力不足、应用效率低的问题,提出一种基于时序特征和自学习机制的检测方法。采用 FT-Transformer(Feature Tokenizer Transformer)架构,结合时序特征分析与 SE-Net(Squeeze-and-Excitation Network)自学习机制,构建 SAFE-Former 检测模型。通过分析 SSH 连接建立过程中密钥和算法协商阶段的关键报文,提取协商报文时延比值或占比特征构建代表性特征,利用特征独立编码机制和 SE-Net 自学习机制捕捉协议交互过程中的时序依赖关系与异常模式,实现对正常与攻击流量的有效区分。实验表明,该方法在互联网环境测试数据集上准确率达 98% 以上,消融实验与多模型对比分析验证了所提特征与模型的有效性。相比现有主流方法,SAFE-Former 在 SSH 中间人攻击检测方面展现出显著优势。

关键词: SSH 协议; 中间人攻击检测; 时序特征; 自学习机制; FT-Transformer

中图分类号: TP393

文献标识码: A

文章编号: 1671-0673(2026)02-0216-07

SAFE-Former: a Method for SSH Man-in-the-Middle Attack Detection

LIU Kun, LIN Wei, JIN Di, LI Ding, CHEN Di

(Information Engineering University, Zhengzhou 450001, China)

Abstract: To address the limitations of traditional methods in secure shell (SSH) man-in-the-middle attack detection, particularly insufficient generalization capability and low application efficiency, a detection method incorporating temporal features and self-learning mechanisms is proposed. In the method, the feature tokenizer transformer (FT-Transformer) architecture is adopted and time series analysis is integrated with squeeze-and-excitation network(SE-Net) self-learning mechanisms to construct the SAFE-Former detection model. Representative features are constructed from delay ratio and proportion characteristics of negotiation packets by analyzing critical packets during the key exchange and algorithm negotiation phases of SSH connection establishment. The feature tokenization mechanism and SE-Net self-learning mechanism are utilized to capture temporal dependencies and anomalous patterns in protocol interactions, enabling effective differentiation between normal and attack traffic. Experimental results demonstrate that the method achieves an accuracy exceeding 98% on Internet environment test datasets. The effectiveness of the proposed features and model is validated through the ablation studies and multi-model comparative analysis. Compared with existing mainstream approaches, SAFE-Former exhibits significant advantages in SSH man-in-the-middle attack detection.

Key words: SSH protocol; man-in-the-middle attack detection; temporal features; self-learning mechanism; FT-Transformer

收稿日期: 2025-12-31; 修回日期: 2026-02-28

基金项目: 国家自然科学基金青年基金(62302520); 国家自然科学基金(62402524)

作者简介: 刘 骥(2002—), 男, 硕士生, 主要研究方向为网络空间安全。E-mail: 17836956210@163.com

中间人攻击^[1](Man-in-the-Middle Attack, MITM)

作为网络攻击的一种,因其隐蔽不易察觉的特点对网络构成极大威胁。针对TCP协议的应用层加密协议,中间人使自己处于信息的发出者和接收者之间并对双方发出的信息进行接收和处理。在通信被加密保护的情况下,攻击者仍可能利用通信协议的潜在漏洞,通过密钥篡改等技术手段破解加密机制,从而实现对加密信息的解密与窃取,完成中间人攻击的全过程^[2]。

基于SSH(Secure Shell)协议的安全远程登录是目前网络中远程系统管理最常用的加密协议之一^[3]。SSH协议因加密通信特性成为远程管理核心工具,但其密钥和算法协商阶段易受中间人攻击。传统检测方法主要依赖密钥指纹验证或宏观流量统计特征,面临3大核心挑战:1)泛化性差,难以应对未知或变种攻击工具;2)特征可区分性弱,中间人行为在协议层面模仿合法交互,导致基于协议基本特征或简单统计的方法失效;3)效率与适应性不足,复杂网络环境下的噪声和抖动使得传统机器学习模型性能显著下降。

针对上述问题,本文提出一种基于协议交互时序特征与自学习机制的SSH中间人攻击检测方法。该方法的核心思想为:剖析SSH握手过程中密钥与算法协商阶段的微观时序变化,提取时延比特征并构建特征向量;进而,基于BERT模型^[4]设计,构建SAFE-Former(Self-Adaptive Feature Enhancement Transformer)模型。该模型采用FT-Transformer架构对每个时延比特征进行独立建模,通过多头自注意力机制捕获特征间的复杂交互关系;引入SE-Net(Squeeze-and-Excitation Network)自学习模块^[5]自适应地学习通道重要性,无需人工预设权重。实验表明,SAFE-Former模型在检测精度、鲁棒性及泛化性上均优于基线深度学习模型。

本文主要工作如下:1)提出基于时序特征的分析方法,通过多尺度建模SSH握手阶段关键数据包时延,构建8维特征向量,在保持高区分度的同时降低特征冗余;2)设计SAFE-Former检测模型,采用列级Transformer架构实现特征独立建模,引入SE-Net模块对编码后的通道维度进行自适应重加权,充分挖掘时延特征判别信息;3)SAFE-Former模型在多个数据集上准确率与F1分数均超过98%,优于现有方法,消融实验验证了特征区分度及模型组件的有效性。

1 相关工作

SSH协议最高版本为SSH v2.0,除特别声明版本外,本文统称其为SSH协议。针对SSH协议的中间人攻击方式分为两种:降级攻击^[6]和基于SSH-MITM工具的攻击^[7]。降级攻击是一种通过将SSH协议从2.0版本降级到1.0版本来实施中间人攻击的方法。攻击者通过降级攻击绕过更强的加密算法和认证机制等安全特性,从而更容易截获和篡改通信内容。SSH-MITM作为一种开源的SSH安全审计中间人攻击工具,其攻击原理主要体现为密钥替换攻击或身份认证伪造攻击。具体过程为:在客户端和服务端进行密钥交换消息传递时,攻击者利用中间人劫持技术将流量重定向到恶意节点,随后利用伪造的公钥对真实SSH服务器的公钥进行替换。此时中间人在客户端和服务端之间建立起双向的SSH会话,进而实现其攻击目的。

当前,基于流量特征分析和统计的检测方法主要以时间特征、指纹特征等作为主要依据。文献[8]基于SSH异常连接检测,利用数据包大小、发送时间及会话时长等流量统计特征,使用蜜网SSH数据训练检测模型。文献[9]针对加密流量中的中间人攻击识别,提取数据包时间间隔和长度作为特征,运用机器学习算法对加密流量中的社交媒体应用进行分类。文献[10]针对中间人攻击的复杂性,从网络流量中提取数据包大小、收发频率等多维特征,通过特征选择保留对分类最重要的特征子集,提高模型对中间人攻击的检测能力。

加密流量分类领域,研究者提出多种深度学习方法以实现精准识别。Hwang等^[11]将数据包头部的文本信息视为定长的句子,将数据包识别的工作等价于对等长句子的识别,并使用长短期记忆网络(Long Short-Term Memory, LSTM)模型进行识别,在Mirai-RGU数据集上对流量恶意性的判断准确率达到97.2%。Manhas等^[12]融合强化学习与深度学习检测网站中间人攻击,分类器通过基于SSL/TLS协议的网站数据集训练,实验结果准确率达99%以上。郭晓军等^[13]提出ETC-IAM-BiLSTM组合模型,通过极端随机树提取流量特征,结合改进注意力机制与双向LSTM处理时序信息,在Kitsune与Ooter数据集上准确率分别达99.98%与99.94%。

然而,传统的基于统计特征或基于深度包检测难以有效应对加密流量中的中间人攻击。当前基于深度学习的方案虽效果明显,但难以应对标注数

据稀缺、攻击种类多变、模型泛化能力弱等问题。

2 SSH中间人攻击检测方法

针对当前检测存在的现实问题,从协议交互时序特征^[14]入手,分析SSH协议在有无中间人攻击状态下的时延差别,通过提取握手过程中的时序特性构建具有强区分度的时延比特征向量,并基于SAFE-Former模型进行特征学习和分类建模,构建出用于检测SSH中间人攻击的分类模型。

2.1 特征选择与提取

攻击者通常在特定的网络位置或针对某些敏感信息执行针对性中间人攻击。由于中间人要转发客户端与服务端之间的通信报文,TCP握手包在有中间人时从客户端到达中间人的时间小于无中间人时从客户端到达服务端的时间,SSH数据响应

时间在有中间人时比无中间人时的响应时间长。

本文特征符号定义如表1所示。基于SSH协议握手数据包分析^[15]及上述对有无中间人攻击情况下的时延分析,选取SSH协议关键报文的时间差计算时延比并构建特征向量,如图1所示。

表1 特征符号定义表

特征符号	特征名称	含义说明
t_0	握手时延比	反映TCP连接的时延特性,用于识别中间人转发带来的额外延迟
t_1	版本协商时延	客户端与服务端版本协商阶段的耗时
t_2	算法协商时延	密钥与加密算法协商阶段的耗时,中间人操作会导致明显延迟
t_3	密钥交换时延	Diffie-Hellman密钥交换阶段的耗时
t_4	数据传输时延	加密数据传输阶段的平均时延

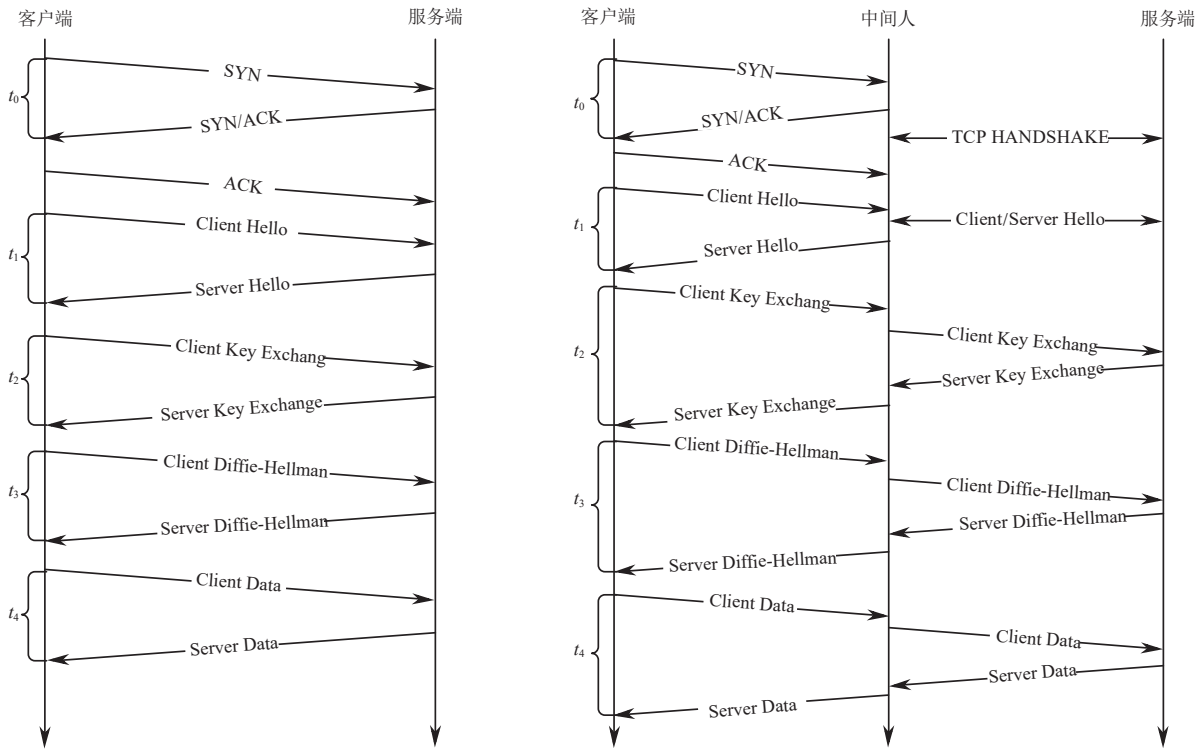


图1 关键报文时间差

假设特征向量 $\mathbf{F} = (r_1, r_2, r_3, \dots, r_n)$, 其中 $r_i = f_i(t_0, t_1, t_2, t_3, t_4)$, n 为特征数量。给出 r_i 计算公式:

$$\begin{aligned} r_1 &= t_2/t_0, r_2 = t_3/t_0, r_3 = t_2/t_1, r_4 = t_3/t_1, r_5 = t_2/t_4, \\ r_6 &= t_3/t_4, r_7 = t_3/t_2, r_8 = (t_1 + t_2 + t_3)/t_0. \end{aligned} \quad (1)$$

式中: r_1 、 r_2 、 r_3 、 r_4 、 r_5 、 r_6 分别为算法协商与密钥交换占各阶段的时延比; r_7 为算法协商与密钥交换的相对时延; r_8 为协商总时延占比。

2.2 模型设计

图2展示了SAFE-Former模型的整体架构,主

要分为4个核心部分:数据预处理层、列级嵌入层、特征增强编码层和多任务输出层。

给定输入特征向量 $\mathbf{F} \in \mathbf{R}^{B \times 8}$, 其中 B 为批次大小, 8 为时延比特征维度。模型首先对输入进行数据归一化预处理, 然后通过列级嵌入将每个特征映射到高维空间, 添加[CLS] token后送入Transformer编码器进行特征交互建模, 经过SE-Net自学习模块增强后提取全局表示, 最后通过分类头完成主分类任务。

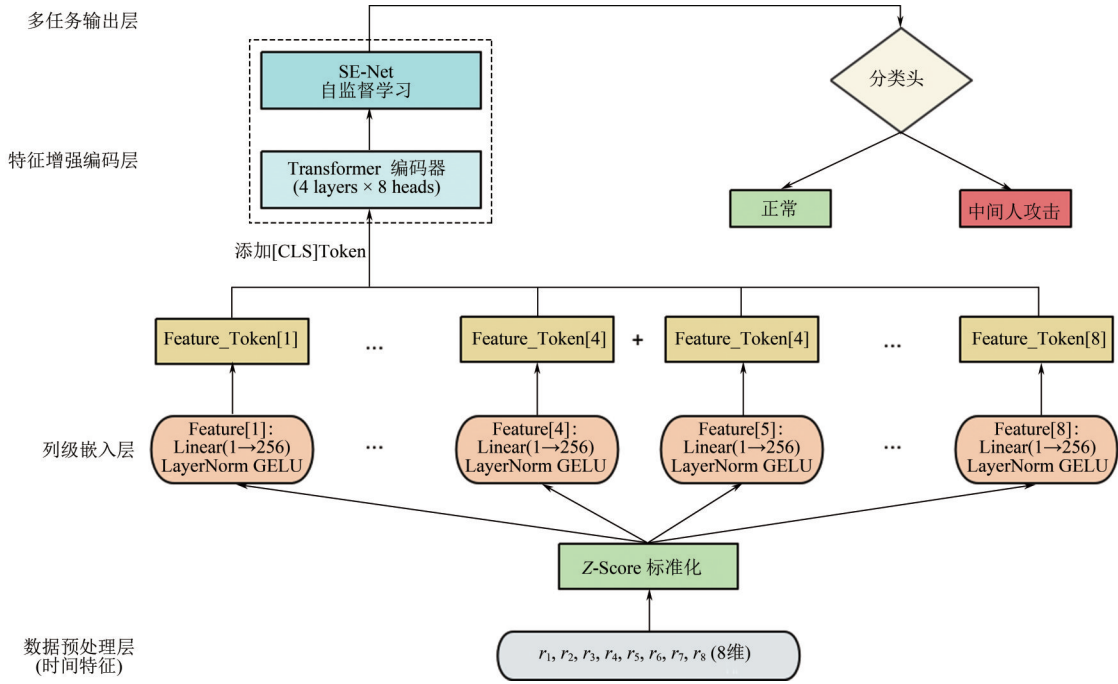


图2 模型整体架构

2.2.1 数据预处理层

数据预处理层主要是对原始时延比进行标准化处理,确保各特征处于相同量级,加速模型收敛。首先,对输入特征向量 $\mathbf{F}$ 中的每个数据进行对数变换以压缩数值范围:

$$r' = \log_2(1 + |r|) \quad (2)$$

随后,对变换后的特征进行Z-score标准化:

$$r'' = (r' - \mu) / \sigma \quad (3)$$

式中: μ 和 σ 分别为训练集特征的均值和标准差。

2.2.2 列级嵌入层

列级嵌入层采用列级嵌入策略,为每个时延比特征设计独立的嵌入网络。列嵌入表示特征的类型属性(如握手时延比、算法协商时延比等)。对于第 i 个特征 $r_i \in \mathbf{R}^{B \times 1}$,通过专用的线性变换和非线性激活进行映射:

$$E_i = \text{GELU}(\text{LN}(\mathbf{W}_i \cdot r_i + \mathbf{b}_i)) \quad (4)$$

式中: $\mathbf{W}_i \in \mathbf{R}^{d \times 1}$ 为第 i 个特征的权重矩阵; $\mathbf{b}_i \in \mathbf{R}^d$ 为第 i 个特征的偏置项; $d=256$ 为模型维度; $\text{LN}(\cdot)$ 为层归一化; $\text{GELU}(\cdot)$ 为高斯误差线性单元激活函数。

为聚合全局信息,在8个特征独立嵌入后,形成特征序列,然后在序列开头添加[CLS] token,通过Transformer编码后能够捕获所有特征的全局表示,用于后续的分类任务。

2.2.3 特征增强编码层

特征增强编码层是SAFE-Former模型的核心,由Transformer编码器和SE-Net自学习模块两部分组成,分别实现列级特征交互建模和通道级自适

应增强。

1)Transformer编码器。编码后的特征向量经过线性投影形成特征序列 $\mathbf{Z} \in \mathbf{R}^{B \times d \times 8}$,其中 B 为批次大小,8为时延比特征维度, $d=256$ 为模型维度。该序列通过标准的多层Transformer编码器进行特征交互建模。每层包含多头自注意力(Multi-head Self-Attention, MSA)和前馈网络(Feed-Forward Network, FFN)两个子层,均采用残差连接和层归一化:

$$\mathbf{Z}' = \text{LN}(\mathbf{Z} + \text{MSA}(\mathbf{Z})) \quad (5)$$

$$\mathbf{H} = \text{LN}(\mathbf{Z}' + \text{FFN}(\mathbf{Z}')) \quad (6)$$

其中多头注意力机制的计算公式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (7)$$

式中: $\mathbf{Q} \in \mathbf{R}^{B \times L \times d_k}$ 为查询矩阵(Query); $\mathbf{K} \in \mathbf{R}^{B \times L \times d_k}$ 为键矩阵(Key); $\mathbf{V} \in \mathbf{R}^{B \times L \times d_v}$ 为值矩阵(Value),3者均由输入序列经独立线性变换映射得到; B 为批次大小; L 为序列长度; d_k 为键向量的维度,本文中, $d_k = \frac{d}{h}$, $d=256$, $h=8$ (注意力头数),故 $d_k=32$; d_v 为值向量维度,通常 $d_v=d_k$ 。

2)SE-Net自学习模块。虽然Transformer编码器通过自注意力机制充分捕获了不同时延特征之间的全局交互关系,但其在特征通道(Channel)维度的注意力分配相对均衡。为增强模型对通道维度的自适应能力,模型在Transformer输出端引入了SE-Net自学习模块。该模块针对列级嵌入后的高维空间,通过学习通道间的依赖关系,实现对不

同时延特征深层表示的自适应增强。该模块通过 Squeeze、Excitation 和 Scale 等 3 个步骤动态调整各通道权重。

首先,模块对 Transformer 输出的序列表示 H 在序列长度维度(包含 8 个时延特征和 1 个[CLS] token,记总长度 $L=9$)上进行全局平均池化,即:

$$z_c = \frac{1}{L} \sum_{l=1}^L H[b, l, :] \quad (8)$$

式中: $H \in \mathbf{R}^{L \times d}$ 为 Transformer 编码器输出的序列表示矩阵; $b \in B$ 为批次维度的索引; l 为序列位置索引, $l=1,2,\dots,L; L=9$; $:$ 表示所有通道; $z_c \in \mathbf{R}^d$ 为全局平均池化后的通道统计向量,即对 L 个位置的表示向量取均值,用于后续 SE-Net 通道权重学习; $c \in d$ 为通道维度索引。

随后,利用降维矩阵 W_D 将通道数压缩并提取核心特征,经 ReLU 激活后,再通过升维矩阵 W_U 恢复至原通道数 d ,最后由 Sigmoid 函数 σ 归一化,生成包含各通道重要性评估的权重向量 $s \in \mathbf{R}^{B \times d}$:

$$s = \sigma(W_U \cdot \text{ReLU}(W_D z_c)) \quad (9)$$

最后,将学习到的通道权重 s 与 Transformer 编码器的原始输出 H 沿着通道维度进行逐元素相乘,得到增强后的特征表示 H' :

$$H' = s \otimes H \quad (10)$$

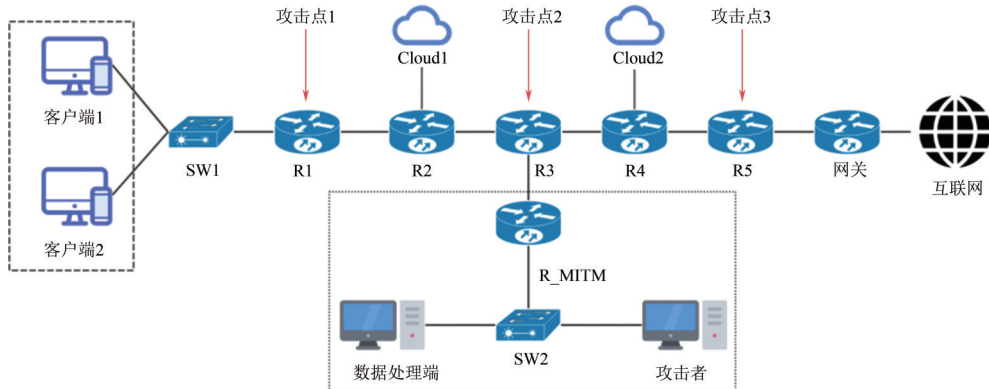


图3 中间人攻击模拟环境拓扑

3.2 数据采集

为使实验数据更具代表性和可靠性,本文分别使用了两台不同区域的云服务器,通过访问云服务器上搭建的SSH服务器采集数据。

数据采集分为正常流量和攻击流量两部分:客户端直接访问两台云服务器SSH服务进行正常流量采集;在R1、R3、R5等3个攻击点及真实互联网环境(Real)下分别实施中间人攻击,进行攻击流量采集。其中真实网络环境的攻击流量通过搭建串接式代理网关服务器采集获取:以一台充当网关路由器的服务器连接互联网,并在其上部署

2.2.4 全局表示提取与多任务输出层

经过特征增强编码后,提取[CLS] token 的表示作为整个输入序列的全局表示,该全局表示融合所有时延比特征的信息,并经过列级交互建模和通道级自适应增强,包含丰富的判别性语义。随后,模型通过多任务输出层,同时完成分类任务。

3 实验设计及结果

为全面评估 SAFE-Former 在 SSH 中间人攻击检测任务中的表现,本节从多个维度展开实验分析。首先,通过与多组基线模型在 SSH 数据集上进行对比实验,验证该模型的性能和效果;其次,通过消融实验探究各模块对整体性能的贡献。

3.1 环境搭建

由于在实际互联网环境中难以实施中间人攻击,本研究基于 EVE-NG 平台搭建模拟攻击环境。同时考虑到中间人位置对检测效果的影响,在模拟环境中共设置 3 个中间人攻击节点(R1、R3、R5),分别对应客户端到服务端路径上的近端、中端和远端位置。利用 SSH-MITM 开源工具实施中间人攻击,通过 Scapy、Wireshark 等工具采集正常与攻击流量。实验环境如图 3 所示。

SSH-MITM 工具,捕获具备真实互联网时延与路由特征的中间人攻击流量。

根据会话五元组对保存的 pcap 文件进行双向会话流拆分,提取 SSH 双向会话流的有效特征,其中正常流量共计 46 656 条,攻击流量条数具体如表 2 所示。

本节针对 SSH 中间人攻击场景构建的自采流量数据集 SSH-MITM-Datasets^[16]已实现开源,包含经预处理得到的 SSH 协议关键报文时延比特征向量文件,同时配套的有生成流程与使用方式的 README 文档。

表2 有效特征记录

攻击点	中间人流量条数
R1	10 323
R3	10 849
R5	10 751
Real	10 904
总计	42 827

3.3 实验与结果分析

3.3.1 评价标准

该实验将正常SSH流量样本看做负类(阴性样本),将中间人攻击的SSH流量样本看做正类(阳性样本),选择准确率 R_{accuracy} 、精确率 $R_{\text{precision}}$ 、召回率 R_{recall} 和 $F1$ 值作为模型训练完成后模型性能评估的评价标准。计算公式为:

$$R_{\text{accuracy}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}} \quad (11)$$

$$R_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (12)$$

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (13)$$

$$F1 = \frac{2R_{\text{precision}}R_{\text{recall}}}{R_{\text{precision}} + R_{\text{recall}}} \quad (14)$$

式中: N_{TP} 表示正确预测为阳性的实例数量; N_{TN} 表示正确预测为阴性的实例数量; N_{FP} 表示错误预测为阳性的实例数量; N_{FN} 表示错误预测为阴性的实例数量。

3.3.2 对比实验

为全面评估本文模型性能,实验选择FS-net^[17]、Deeppacket^[18]、TSCRnn^[19]这3个典型加密流量分类模型作为基线模型。所有基线模型均使用同一特征数据集进行训练与测试,超参数均经过优化。

数据集选择方面,本实验数据从采集的SSH流量中随机抽取正常流量40 000条,从R1、R3、R5及真实网络环境(Real)的攻击流量中各抽取10 000条,共计80 000条样本,按攻击节点将数据等分为4个数据集:DSP1、DSP2、DSP3、DSP_Real。其中DSP1数据集用于深度学习模型的训练,按照8:2的比例分为训练集和验证集。其余3个数据集用于对模型进行全面的泛化性测试和性能评估。表3展示了在不同模型下的测试结果。

表3 对比实验结果

模型	DSP2				DSP3				DSP_Real			
	准确率	精确率	召回率	$F1$ 值	准确率	精确率	召回率	$F1$ 值	准确率	精确率	召回率	$F1$ 值
FS-net	0.971 2	0.973 5	0.968 9	0.971 2	0.968 3	0.970 5	0.966 2	0.968 4	0.956 2	0.968 1	0.962 3	0.956 1
Deeppacket	0.978 6	0.980 2	0.977 1	0.978 6	0.972 5	0.975 5	0.978 6	0.972 5	0.966 8	0.948 2	0.925 5	0.936 4
TSCRnn	0.984 5	0.986 1	0.983 2	0.984 5	0.979 3	0.981 4	0.977 5	0.979 1	0.965 1	0.955 5	0.931 0	0.932 6
SAFE-Former	0.997 0	0.998 3	0.994 5	0.996 9	0.986 2	0.997 5	0.972 7	0.986 1	0.981 7	0.974 4	0.985 4	0.984 8

由表3结果可得,SAFE-Former模型在模拟环境和物理环境流量数据集上均取得了较为优异的性能,准确率和 $F1$ 分数均超过98%,整体表现优于其他基线模型。相比之下,FS-Net依赖于传统统计特征,对时序细微变化的捕捉不足,导致其性能相对较低。Deeppacket对数据包载荷特征的依赖在SSH全加密环境下受限,且对网络时延抖动更为敏感。TSCRnn虽专为时序建模设计,能较好捕捉序列依赖关系,但SAFE-Former通过分窗口时序特征统计策略精细刻画握手阶段时延变化规律,结合特征增强编码与自注意力机制获得了更丰富的判别信息,性能优于单纯的端到端时序模型。

3.3.3 不同时序特征对模型性能的影响

为验证所提出的时延比特特征的可区分性和有效性,本节设置时序特征的消融实验。实验设计基于SAFE-Former模型,在完整训练的模型基础上,逐个将某一维特征在测试阶段置零(即在输入

模型前将该特征值设为0),观察模型性能的下降幅度。实验结果如图4所示,图中橙色柱代表移除特定特征(r_2 或 r_7)后的实验结果。

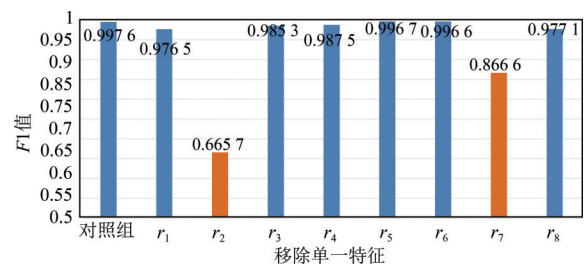


图4 单一特征消融实验结果

从图4的实验结果分析得出:特征 r_2 和特征 r_7 在检测中起到主导作用,移除后 $F1$ 值分别相对下降约33.27%和13.13%,这是由于中间人在密钥交换阶段需分别与客户端和服务端完成双重协商,引入的额外时延构成了核心判别依据。其余特征虽单独贡献较小,但能捕获握手过程不同阶段的

时序关系,在网络抖动或异常流量干扰下有效提升模型鲁棒性。

3.3.4 不同模型模块对检测性能的影响

为分析 SAFE-Former 模型中各组件对性能的贡献,设计 3 组消融实验及 1 组对照实验。具体方法为:以 SAFE-Former 完整模型为基准,在相同的

数据集和训练配置下,分别设置移除模型的 SE-Net 自学习模块(w/o SE-Net)、移除独立列级嵌入模块(w/o ColumnEmbedding)以及移除 Transformer 模块,仅保留列级嵌入层与 SE-Net 自学习模块(w/o Transformer)3 组消融实验,对比不同条件下模型在各个数据集下的表现,实验结果如图 5 所示。

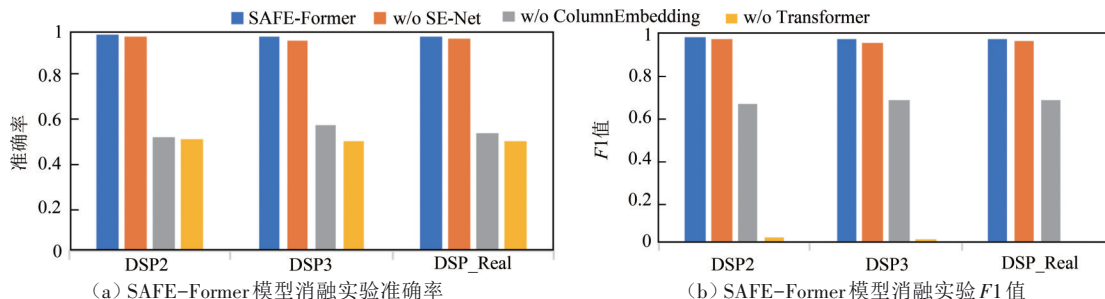


图 5 SAFE-Former 模型消融实验结果

实验结果表明,各模块对模型性能的贡献存在显著差异。移除 SE-Net 模块导致性能下降约 1.5%,表明自适应特征加权机制能够稳定提升模型性能。去除独立列级嵌入层后,准确率和 F1 值分别从接近 1.0 骤降至 0.55 和 0.68,说明特征独立建模对捕捉特征特异性至关重要。而移除 Transformer 模块则导致性能崩溃性下降,准确率接近随机水平,F1 值趋近于 0,验证了 Transformer 作为核心组件在建模特征间复杂交互关系中不可替代的作用。

4 结束语

本文通过分析 SSH 协议握手协商阶段的关键报文,提出一种基于时序特征和自学习机制的 SAFE-Former 检测模型,用于 SSH 中间人攻击检测,通过提取协商报文时延比值或占比特征构建特征向量数据集,使用 SAFE-Former 模型训练得到最优分类模型。对于 SSH 协议,所选特征具有较高的可区分性,同时所提方法能够克服中间人在不同网络位置的影响,在不同时延模拟攻击的环境及实网环境采集的流量样本中达到了 98% 以上的准确率。在实际网络环境中,可以将检测模型部署应用在网关等关键节点,增强网络边界防御的能力。未来将致力于探索模型在复杂网络拓扑下的检测性能与实时性,并进一步提升 SAFE-Former 模型的泛化性与鲁棒性,探索其在 HTTPS 等更多加密协议环境下的中间人攻击检测效能。

参考文献:

[1] CONTI M, DRAGONI N, LESYK V. A survey of man in

the middle attacks[J]. IEEE Communications Surveys & Tutorials, 2016, 18(3):2027-2051.

- [2] MALLIK A, AHSAN A, SHAHADAT M M Z, et al. Understanding man-in-the-middle-attack through survey of literature[J]. Indonesian Journal of Computing, Engineering and Design, 2019, 1(1):44.
- [3] 张杰.SSH 协议的研究与实现[D].合肥:中国科学技术大学,2002:11-13.
- [4] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minnesota, USA: Association for Computational Linguistics, 2019:4171-4186.
- [5] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018:7132-7141.
- [6] 熊克琦.SSH 协议的中间人攻击研究[D].北京:北京邮电大学,2008:35-40.
- [7] KAISER M. SSH-MITM [EB/OL]. (2021-01-25) [2025-12-15]. <https://github.com/ssh-mitm/ssh-mitm>.
- [8] GONZÁLEZ S, HERRERO Á, SEDANO J, et al. Different approaches for the detection of SSH anomalous connections [J]. Logic Journal of the IGPL, 2016, 24(1): 104-114.
- [9] AL-HABABI A, TOKGOZ S C. Man-in-the-middle attacks to detect and identify services in encrypted network flows using machine learning[C]//Proceedings of the 3rd International Conference on Advanced Communication Technologies and Networking. Piscataway, USA: IEEE, 2020: 1-5.
- [10] KUMAR A, SHARMA I, MITTAL S, et al. Enhancing se

(下转第 230 页)