# 融合多尺度语义与VMD-BiLSTM的恶意APP检测模型

许国良，时磊，邱思琦，许宇
（重庆邮电大学通信与信息工程学院，重庆 400065）

**摘　要：**针对加密流量中恶意APP特征难提取、背景噪声大及模型缺乏透明度的问题，提出一种一种融合变分模态分解(VMD)、Attention-BiLSTM与SHAP机制的检测模型。首先，针对移动APP流量的多尺度特性，设计自适应多尺度窗口机制，动态提取并构建融合语义的高维时间序列；其次，为处理该结构化序列中交织的复杂环境噪声，引入变分模态分解(VMD)进行频域平稳化降噪，并利用结合焦点损失的Attention-BiLSTM网络精准捕获长程时序依赖；最后，引入SHAP机制量化特征的边际贡献，提供事后归因解释以辅助决策溯源。实验表明，该模型准确率达98.18%，在实现较高精度检测的同时，提升了模型判决的透明度与可信度。

**关键词：**恶意APP识别；自适应多尺度窗口；流量异常检测；VMD；BiLSTM；可解释性分析
**中图分类号：**TP393.08
**文献标志码：**A

# Malicious APP detection model integrating multi-scale semantics and VMD-BiLSTM

Xu Guoliang, Shi Lei, Qiu Siqi, Xu Yu
School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065，China

**Abstract:** To address the challenges of difficult feature extraction, high background noise, and a lack of model transparency in identifying malicious applications within encrypted traffic, this paper proposes a novel detection model integrating Variational Mode Decomposition (VMD), Attention-BiLSTM, and the SHAP mechanism. First, targeting the multi-scale characteristics of mobile APP traffic, an adaptive multi-scale window mechanism is designed to dynamically extract and construct semantic-driven high-dimensional time series. Second, to mitigate the complex environmental noise intertwined within these structured sequences, VMD is introduced for frequency-domain stationary denoising. Subsequently, an Attention-BiLSTM network coupled with Focal Loss is employed to accurately capture long-range temporal dependencies. Finally, the SHAP mechanism is incorporated to quantify the marginal contributions of features, providing post-hoc attribution explanations to facilitate decision traceability. Experimental results demonstrated that the proposed model achieved an accuracy of 98.18%, successfully realizing high-precision detection while simultaneously enhancing the transparency and credibility of the model's decision-making process.

**Keywords:** malicious APP identification, adaptive multi-scale window, traffic anomaly detection, VMD, BiLSTM, interpretability analysis

## 0　引言

随着移动互联网与智能终端的深度普及，移动应用程序(Application, APP)已渗透至金融支付、社交通信、物联网控制等关键领域，但恶意APP带来的安全风险日益突出——其通过隐藏恶意行为、伪装正常通信等方式，窃取用户信息、破坏设备功能、引发财产损失，对移动网络空间安全构成严重威胁。

当前基于流量异常检测的恶意APP识别面临

收稿日期：XXXX–XX–XX；修回日期：XXXX–XX–XX
通信作者：许国良，xugl@cqupt.edu.cn

多尺度特征利用难、背景流量非平稳以及深度学习模型缺乏可解释性等核心挑战。首先是传统方法难以兼顾识别精度、泛化能力与场景适配性。移动恶意行为在时间维度上具有天然的多尺度特性，而传统依赖固定长度滑动窗口的“单尺度”检测方法在处理复杂混合流量时存在显著缺陷，极易产生“尺度失配”，导致模型难以兼顾瞬时流量激增与潜伏性慢变趋势的精准捕捉；其次，移动APP在复杂的无线网络环境下运行，其产生的动态流数据具有极强的非平稳性与模式不确定性，原始流量中交织的冗余环境噪声极易导致模型感知混乱，难以提取到稳定的判别基准；最后，随着全报文加密技术的广泛普及，传统依赖载荷内容的检测逻辑失效，单一的流统计特征难以还原APP真实的业务语义；同时，当前的深度学习检测模型在输出判决结果时往往被视为缺乏物理意义的“黑盒”，难以实现对异常判决背后的具体恶意行为模式进行归因与溯源，这限制了其在网络安全防御体系中的透明度与可靠性。

## 1 研究背景

### 1.1 国内外研究现状

研究发现，正常流量和异常流量的时变信号在频率上存在较大差异，将网络流量视为信号并利用频域特征分离多尺度异常成为该领域的重要方向。Brynielsson等[1]对网络流量进行离散傅里叶变换后，利用频谱分析方法评估了不同攻击场景下的检测效果。由于大规模网络流量具有宏观信号的特征，小波分析被广泛应用于检测短期和长期网络流量异常。例如，Cheng等[2]利用离散小波变换将原始流量变换为不同频域下的数据序列，为异常检测提供了多样性信息；Wang等[3]将小波频率分析嵌入深度学习框架，提取频域特征进行无监督的异常检测；Fouladi等[4]则提出了基于离散小波变换和自编码器(Auto-Encoder, AE)神经网络的分布式拒绝服务攻击检测方案，取得了较好效果。

利用深度学习挖掘大规模复杂网络流量中的异常是当前研究的热点，相关研究按模型结构主要分为单结构模型与多结构模型。在单结构模型方面，Albahar[5]基于循环神经网络模型实现了对入侵异常的有效检测；Pei等[6]提出了一种基于长短期记忆网络(Long Short-Term Memory, LSTM)结构自编码的网络流量异常检测方法；Zong等[5]设计的深度自编码高斯混合模型将AE与高斯混合模型结合，实现了数据降维与密度估计的同步优化。此外，生成式模型也被广泛应用，如Geiger[8]等提出的TadGAN利用双向生成对抗网络(Generative Adversarial Network, GAN)实现时间序列异常检测，Patil等[9]提出的PCA-BiGAN利用主成分分析结合双向GAN提高了检测效率，邹福泰等[10]则利用GAN生成扩充了恶意流量特征样本。

随着深度学习的演进，多结构组合模型逐渐兴起以应对更复杂的特征关系。Chen等[11]提出的DAEMON模型结合了变分自编码器与GAN，增强了对异构数据的适应性；麻文刚等[12]使用堆叠LSTM结合残差网络提取不同深度的流量特征，改善了梯度消失问题；Chouhan等[13]利用堆叠自编码器(Stacked Auto-Encoder, SAE)与深度卷积神经网络(Convolutional Neural Network, CNN)构建了CBR-CNN模型；Yang[14]将堆叠自编码器与LSTM结合建模有效特征的时间结构；Ullah等[15]则混合CNN与BiLSTM等循环神经网络实现了轻量级的异常流量检测。

## 2 时间序列构建及特征提取

本节作为检测模型的前端特征工程模块，旨在通过分阶段的流量语义分类与自适应时间尺度划分，将原始离散、非平稳的网络请求转化为适用于后续融合变分模态分解、注意力机制的双向长短期记忆网络深度建模的结构化多维时序特征矩阵。

### 2.1 基于标准化威胁框架的流量语义类型识别

本节通过分阶段流量类型识别，实现流量语义分类，为后续高频语义和统计特征分离建模奠定基础：先基于关键词规则完成请求初步分类，再对“其他”类流量引入结构模板与聚类分析细化识别，提升语义分析的完整性与区分度。

#### 2.1.1 基于OWASP的威胁语义初步分类

在真实的网络对抗环境中，现代移动恶意软件(如银行木马、勒索软件、间谍软件等)在网络层的核心生命周期内展现出高度趋同的恶意行为模式。相关学术研究表明，绝大多数移动恶意软件严重依赖网络接口来协调操作、窃取用户隐私数据并发起攻击活动[16]。具体而言，恶意软件最高频的网络行为特征包括：利用伪造接口或劫持手段窃取用户

凭证与敏感信息；通过应用层协议伪装成正常的Web流量以隐藏其命令与控制通信；以及在运行时动态拉取外部恶意载荷以规避应用商店的静态审查。此外，将受害者的敏感信息)偷偷回传至远程恶意服务器也是其标志性行为[17]。

为了准确且科学地捕捉上述底层攻击意图，构建了基于OWASP Mobile Top 10的威胁语义映射体系。该体系将流量特征的提取直接锚定在国际公认的安全威胁分类基准与学术界公认的高频恶意行为模式上，将请求类型集合定义为$C=\{c_1$(凭证与验证风险), $c_2$(供应链与动态加载风险), $c_3$(恶意载荷注入与校验缺失风险)，$c_4$(其他风险)$\}$：

1)凭证与验证风险：对应OWASP M1: Improper Credential Usage与M3: Insecure Authentication，凭证劫持与窃密是移动恶意软件最核心的变现手段。恶意软件通常通过伪造登录接口或拦截凭证传输获取用户敏感访问权限。

2)供应链与动态加载风险(对应OWASP M2: Inadequate Supply Chain Security)：此分类直接映射OWASP框架中的供应链安全威胁，用于监控异常的第三方资源加载与未知代码注入过程。

3)恶意载荷注入与校验缺失风险(对应OWASP M4: Insufficient Input/Output Validation)：移动应用在与服务端交互或处理深层链接时，若缺乏严格的输入/输出校验，攻击者极易通过URL参数注入恶意载荷。

每一类$c_j$对应一组依据OWASP安全知识库与前沿漏洞分析提取的威胁特征关键词集合$K_j$。定义匹配函数，当一个请求在多重匹配结果中时，通过基于安全威胁严重程度的类别优先级函数$pi(c_j)$确定唯一分类标签。同时，引入路径特征识别规则辅助判定API请求，作为后续威胁语义建模的补充特征。

#### 2.1.2　语义不明确的流量的结构细化识别

针对关键词缺失无法明确分类其语义的其他类请求，设计基于路径-参数模板的结构建模方法：首先通过URL标准化函数构造结构模板(保留路径与参数结构，排除参数值与冗余前缀)；然后构建归一化编辑距离矩阵量化模板相似性；最后采用DBSCAN算法聚类，对同一聚类簇请求赋予“other_1”“other_2”等编号标签，无法归类的标记为“other_unclassified”，弥补初步识别在未知风险请求处理上的短板。考虑到分类标签可能过多，细化的分类标签只在计算后续行为变化特征时进行使用。

### 2.2　自适应滑动窗口场景识别

本文所引入的场景语义-时间尺度适配机制依据场景类型动态分配时间粒度，短时场景用小窗口保留高频细节，长时场景用大窗口捕捉趋势变化。在此基础上，设计自适应滑动窗口方法识别高层次行为场景，动态调整窗口策略适应不同场景时间尺度，增强异常检测的时序感知能力。

#### 2.2.1　自适应窗口构造机制

设请求数据按时间升序排列为$R=\{r_1,r_2,\cdots,r_n\}$，其中每个请$r_i$包含时间戳$t_i$及其对应的已识别请求类别$c_i$。窗口构造以时间滑动方式进行，起始于当前请求的时间$t_s$，以初始窗口长度$\Delta_0$向后滑动，形成候选窗口区间$[t_s,t_s+\Delta]$，对应的窗口请求集合定义为

$$W(t_s,\Delta)=\{\, r_i\in R \,|\, t_i\in[t_s,t_s+\Delta)\,\} \tag{1}$$

为保证窗口内具有足够统计样本以实现稳健的场景识别，若$|W(t_s,\Delta)|<\theta$，则动态扩大窗口长度$\Delta\leftarrow\Delta+\epsilon$，直到满足$|W(t_s,\Delta)|\geqslant\theta$。其中$\theta$为最小请求数阈值，$\epsilon$为扩展步长。

#### 2.2.2　场景识别函数设计

对窗口内所有请求类别集合$\{c_1,c_2,\cdots,c_m\}$，统计各类请求频数

$$f_j=\mathrm{count}(c_j) \tag{2}$$

设总请求数为$m$,定义场景分类函数$\Psi(W)$

$$\Psi(W)=\begin{cases} \textit{凭证与验证风险场景} & \text{if } f_{c_1}>0 \\ \textit{隐私控制与数据外泄场景} & \text{else if } \dfrac{f_{c_2}}{m}>0.4 \\ \textit{恶意载荷注入与校验缺失场景} & \text{else if } f_{c_3}>0.4 \\ \textit{其他} & \textit{otherwise} \end{cases} \tag{3}$$

#### 2.2.3　场景驱动的窗口细化策略

由于不同场景对应的时间尺度差异显著，本文进一步引入场景驱动的窗口重构策略：针对已识别场景$\Psi(W)$，采用差异化规则对窗口进行细化拆分；拆分后的每个子窗口均继承原场景标签并记录独立编号，实现行为片段的局部精准标注。最终，数据集中每一条请求均被赋予所属子窗口编号与场景类

别，完成基于流量序列的结构化语义分段。

### 2.3　时间序列构建

将离散请求事件转化为结构化时间序列，为后续模型提供高维、时间敏感的输入基础。以时间窗口为单位，从四个维度提取语义和统计特征，按时间顺序组织形成序列数据。

设窗口 $W_{ij}$ 表示第 $i$ 个场景 $s_i$ 下的第 $j$ 个时间窗口，窗口起始时间为 $t_{ij}$，请求集合为 $\{r_1,r_2,\cdots,r_n\}$，则特征设计如下

**1)** 通信频率特征

核心指标为API请求频率 $f_{\text{API}}$，计算公式为

$$f_{\text{API}}=\frac{N_{\text{API}}}{T_{\text{win}}} \tag{4}$$

其中 $N_{\text{API}}$ 为窗口内标记为API请求的数量，$T_{\text{win}}$ 为窗口时间跨度(秒)。

2) 负载分布特征

包含两类关键指标：一是请求体积均值 $\mu_s\mu_s$ 与方差 $\sigma_s^2$，计算公式分别为

$$\mu_s=\frac{1}{n}\sum_{k=1}^{n}s_k,\sigma_s^2=\frac{1}{n}\sum_{k=1}^{n}(s_k-\mu_s)^2 \tag{5}$$

其中 $s_k$ 为第 $k$ 个请求的请求大小。

二是响应时间均值 $\mu_r$ 与方差 $\sigma_r^2$，计算公式分别为

$$\mu_r=\frac{1}{n}\sum_{k=1}^{n}r_k,\sigma_r^2=\frac{1}{n}\sum_{k=1}^{n}(r_k-\mu_r)^2 \tag{6}$$

其中 $r_k$ 为第 $k$ 个请求的响应均值。

上述指标描述了每个窗口内的通信负载与服务响应效率，异常窗口可能出现异常大的方差或平均值。

3) 语义类型分布特征

重点关注不同安全威胁类别在当前时间窗口内的请求占比。基于前文2.1节构建的OWASP移动威胁分类集合 $C$ 进行计算，设某类风险请求 $c_k$ 在窗口 $W_{ij}$ 中的占比为 $p_{c_k}$，其计算公式为

$$p_{c_k}=\frac{N_{c_k}}{n}\ (k\in\{1,2,3,4\}) \tag{7}$$

其中 $N_{c_k}$ 表示该时间窗口中被分类优先级函数或聚类算法标记为 $c_k$ 类的请求数量，$n$ 为窗口内的总请求数。

**4)** 行为变化特征

核心指标为请求类型切换频率 $f_{\text{switch}}$

$$f_{\text{switch}}=\frac{\sum_{k=2}^{n}I(c_k\neq c_{k-1})}{n} \tag{8}$$

其中 $I$ 为指示函数。

综合上述特征设计，对每个窗口提取如下12维特征向量

$$\mathrm{x}_{ij}=[f_{\text{API}},\mu_s,\sigma_s^2,\mu_r,\sigma_r^2,p_{c_1},p_{c_2},p_{c_3},p_{c_4},f_{\text{switch}}] \tag{9}$$

最终所有特征向量按时间戳 $t_{ij}$ 升序排列，构成完整的全局时间序列

$$X=[x_1,x_2,\cdots,x_T] \tag{10}$$

构建后的时间序列通过窗口级建模保留了行为的时间依赖关系，使用特征压缩兼顾语义类型与通信特征，可灵活适用于不同时间粒度下的检测任务，可直接作为后续异常检测模型或传统聚类算法的输入，实现对时间段级别的行为预测与异常识别。

## 3　基于 VMD 与 Attention-BiLSTM 的深度检测模型

在完成多尺度时间序列构建后，为解决真实网络流量非平稳性导致的“感知混乱”以及深度学习模型的“黑盒”问题，本文设计了融合变分模态分解、注意力机制的双向长短期记忆网络以及自适应焦点损失的深度检测模型，并引入SHAP理论实现判决归因。

### 3.1　方法框架设计

针对当前移动APP网络流数据动态性强、多尺度异常特征难以有效提取，以及深度学习模型在复杂加密环境下缺乏可解释性等挑战，本文提出并实现了一种融合多尺度语义与变分模态分解、注意力机制的双向长短期记忆网络(VMD-BiLSTM)的恶意APP检测模型，整体框架如图1所示。

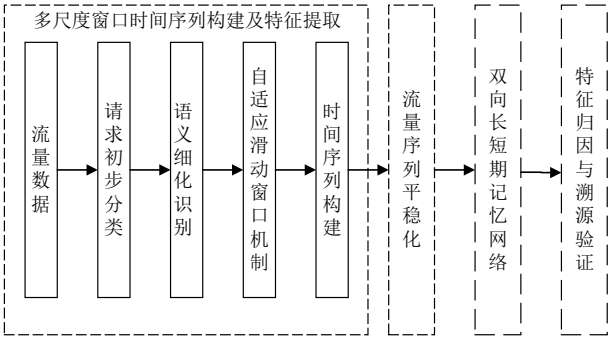



图1　融合多尺度语义与VMD-BiLSTM的恶意APP检测模型框架

在特征构建与预处理阶段，模型首先执行多尺度窗口时间序列的构建与特征提取。系统结合领域知识关键词库与基于路径-参数模板的DBSCAN聚类技术，实现对显式与隐式流量场景的动态识别，并自适应地为短时高频场景分配小窗口、为长时稳定性场景分配大窗口，进而提取通信频率、负载分布及语义类型等多维特征以构建结构化时间序列。随后，针对真实网络流量极强的非平稳性，引入变分模态分解(Variational Mode Decomposition, VMD)算法对高噪特征进行频域内的迭代寻优与平稳化处理，VMD将复杂的时序特征自适应解耦为多个具有物理意义且互不混叠的本征模态函数(IMF)，在剔除高频无序噪声后，重构出高质量、低噪的平稳特征矩阵，为后续深度学习提供可靠的判别基准。

在深度判决与溯源解释阶段，本文设计了融合注意力机制与自适应焦点损失(Adaptive Focal Loss, AFL)的BiLSTM深度判决网络。该网络利用BiLSTM充分捕获流量特征在双向时间轴上的长程演化依赖，并通过注意力机制动态加权，强制模型聚焦于关键的恶意行为突变点。同时，在训练阶段引入AFL机制动态调节损失权重，有效克服了真实网络环境中样本极度不平衡带来的分类偏差难题。最后，为打破深度学习的“黑盒”特性并形成完整的防御闭环，模型引入了沙普利加和解释(SHapley Additive exPlanations, SHAP)框架。该机制通过量化全局特征重要性与局部实例的边际贡献，直观揭示了模型在特定时间窗口内基于关键流量波动做出判定的物理逻辑，实现了从底层特征到高层安全语义的可视化溯源。

### 3.2　基于VMD的流量序列平稳化解耦

真实移动网络流量中广泛交织着正常的并发业务与环境延迟抖动，具有极强的非平稳性与模式不确定性，若直接将原始特征序列输入深度学习模型，冗余背景噪声极易干扰判别基准的提取。因此，本节引入VMD算法对前端构建的多维时间序列$X=[x_1, x_2,\cdots,x_T]$进行频域平稳化处理。VMD的理论核心在于构建并求解一个受约束的变分问题，旨在将复杂信号自适应地解耦为预设数量$K$的本征模态函数。其变分约束模型可表示为

$$\min_{\{u_k\},\{\omega_k\}}\left\{\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\} \quad (11)$$

$$\text{subject to}\ \sum_{k=1}^{K}u_k(t)=f(t)$$

其中，$u_k(t)$为第$k$个模态分量，$\omega_k$为其中心频率，$\delta(t)$为狄拉克分布。模型利用交替方向乘子法在频域内进行迭代寻优，将复杂交织的短时突发调用与长时潜伏流量严格解耦为$K$个具有物理意义且互不混叠的IMF分量，在剔除代表无序噪声的高频残余分量后，系统将保留的平稳IMF分量按序拼接，重构为低噪、平稳的多维时序特征矩阵，作为下一阶段的统一输入。

### 3.3　融合Attention机制与AFL的BiLSTM网络

在获取平稳的多维IMF特征矩阵后，本文构建了Attention-BiLSTM深度网络，以精准捕获隐蔽恶意行为的长程时序依赖。

#### 3.3.1　双向时序状态提取

重构的特征矩阵首先进入BiLSTM网络。通过内部的遗忘门、输入门和输出门协同计算，BiLSTM同时在过去与未来双向时间轴上对流量演化特征进行记忆与提取，输出包含长程上下文信息的隐藏状态序列$H=\{h_1, h_2, ..., h_T\}$。

#### 3.3.2　动态注意力特征聚焦

针对大量常规心跳流量易掩盖稀疏恶意调用的问题，模型将隐藏状态序列$H$输入Attention层。该层通过可学习的参数矩阵计算各时间步的对齐得分，并经Softmax函数归一化生成动态注意力权重分布。该机制强制网络为“异常大体积传输”或“敏感接口高频切换”等关键恶意时间步分配更高权重，最终加权求和生成高度浓缩的上下文向量。

#### 3.3.3　基于AFL的分类失衡优化

在实际流量检测中，良性样本与恶意样本数量往往存在严重失衡，传统重采样方法易破坏流量的原始时序依赖。因此，本文在反向传播阶段引入自适应焦点损失。AFL通过引入动态调制因子$(1-p_t)^{\gamma}$，对于高置信度的易分样本(常规良性流量)，其损失贡献被大幅衰减；而对于低置信度的难分隐蔽恶意流量，该因子保持较高水平。这一机制强行引导网络优化器聚焦于少数类样本边界，在不改变数据物理分布的前提下实现了模型参数的稳健拟合。

### 3.4　基于SHAP机制的判决溯源与解释

为打破深度网络作为高维分类器的“黑盒”特性并形成完整的安全防御闭环，模型在测试评估阶段引入了SHAP框架。基于协同博弈论的SHAP机制将模型预测视为特征间的博弈，通过计算各特征分量在所有特征子集组合中的预测增量并加权平均，求得反映特征边际贡献的沙普利值。该模块包含两个维度的归因：

全局特征归因：量化各语义特征与时序模态在整体分类决策中的边际贡献，验证特定高频语义在判定特定恶意家族时的核心主导作用。

局部实例溯源：针对单一被判定为恶意的高危告警样本，SHAP精确计算并在特定时间窗口内可视化特征波动对结果的驱动拉力，清晰展示模型捕获异常波动的判定逻辑，为安全人员提供从“底层数据”到“上层语义”的完整检测证据链。

## 4　实验与结果分析

### 4.1　实验数据集

本文实验使用的数据集来源于加拿大通信安全研究中心于2020年发布的CICMalDroid2020数据集。该数据集是专为移动恶意行为检测研究构建的大规模Android应用行为集，涵盖良性应用与多种恶意样本，具有真实性强、行为多样、网络流量可观测等优点，广泛应用于Android安全分析与行为建模等研究领域。

**表1**　　**数据集详情**

| APP类别 | 所属家族 | 数量 |
|---|---|---|
| Malware | Adware | 1253 |
| | Banking | 2100 |
| | SMS | 3904 |
| | Riskware | 2546 |
| Benign | Benign | 1795 |

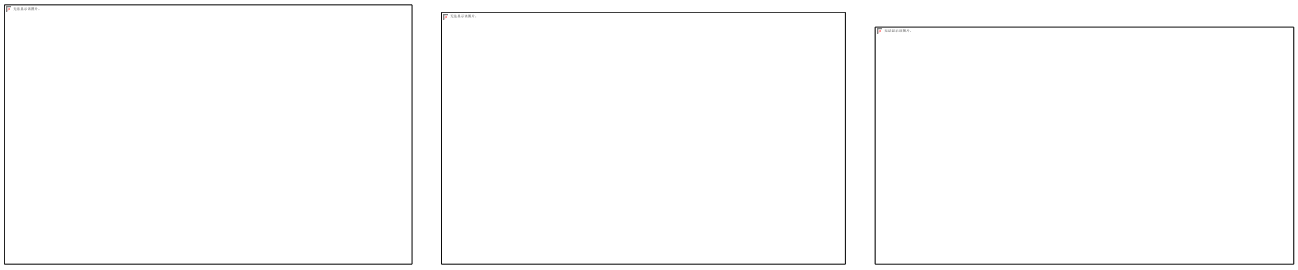
为直观验证该数据集在流量特征上的区分度，对数据集中典型样本的流量特征进行可视化分析，某良性应用的流量特征如图2所示，某恶意应用的流量特征如图3所示。良性应用请求数量稳定、请求大小与响应时间分布均匀，契合正常流量的统计稳定性；恶意应用则出现请求数量骤降、请求大小异常峰值等特征。

### 4.2　实验平台

实验使用Windows操作系统，CPU型号Intel i7-13700KF，GPU型号NVIDIA GeForce GTX 4090Ti显卡，显存32G。

### 4.3　全局效能与分类阈值寻优

将多尺度时间序列输入VMD-BiLSTM模型进



(a)　窗口请求数量变化图　　(b)　请求大小分布图　　(c)　请求大小vs响应时间KDE图

图2　某良性应用流量特征

(a)　窗口请求数量变化图　　(b)　请求大小分布图　　(c)　请求大小vs响应时间KDE图

图3　某恶意应用流量特征

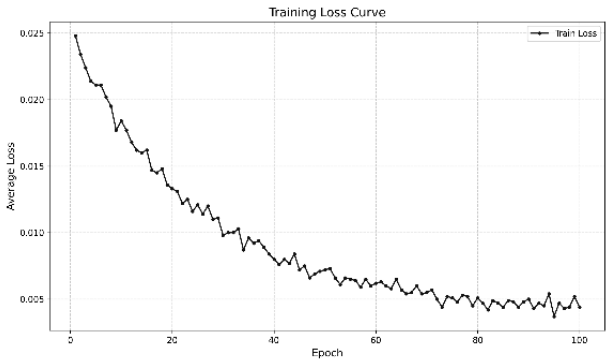



图4　训练损失曲线图

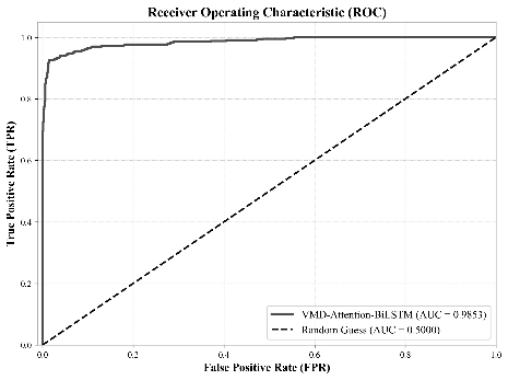



图5　受试者工作特征(Receiver Operating Character Curve, ROC)曲线图

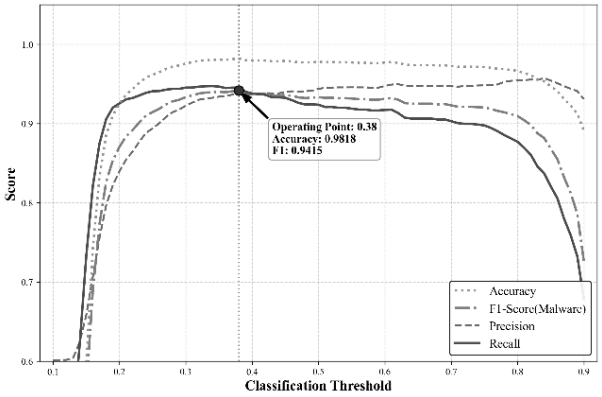



图6　恶意类敏感度分析图

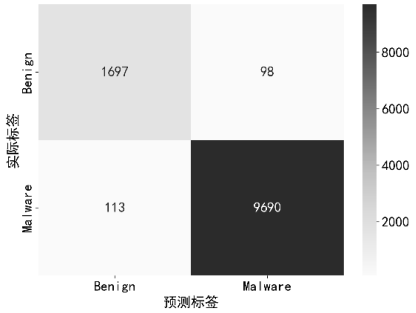



图7　混淆矩阵

行训练。如图4所示，模型平均损失在约60个Epoch后趋于平缓，最终稳定在0.005的极低水平，验证了模型在处理大规模高噪特征时的参数稳健性。在全局区分能力评估中，如图5所示，模型的ROC紧贴坐标系左上角，曲线下面积高达0.9853。

针对类别不平衡问题，本文引入了分类阈值敏感度寻优机制。如图6所示，当分类阈值下调至0.38时，精确率与召回率达到帕累托最优平衡，恶意类F1-Score达到全局峰值0.9415，全局准确率达到0.9818。

如图7所示，混淆矩阵直观地量化了模型在测试集上的多维分类表现。具体而言，模型成功精准拦截了9690个恶意样本(真阳性，TP)，并正确放行了1697个良性样本(真阴性，TN)。在误判控制方面，仅有98个良性网络流被误报为恶意行为(假阳性，FP)，113个恶意流被错判为良性(假阴性，FN)。

### 4.4　模型可解释性分析

为了进一步验证模型判决的可靠性并提供直观的防御证据，本模型通过引入时间注意力机制与SHAP归因理论，从宏观趋势到微观实例对识别模型进行深度透视。

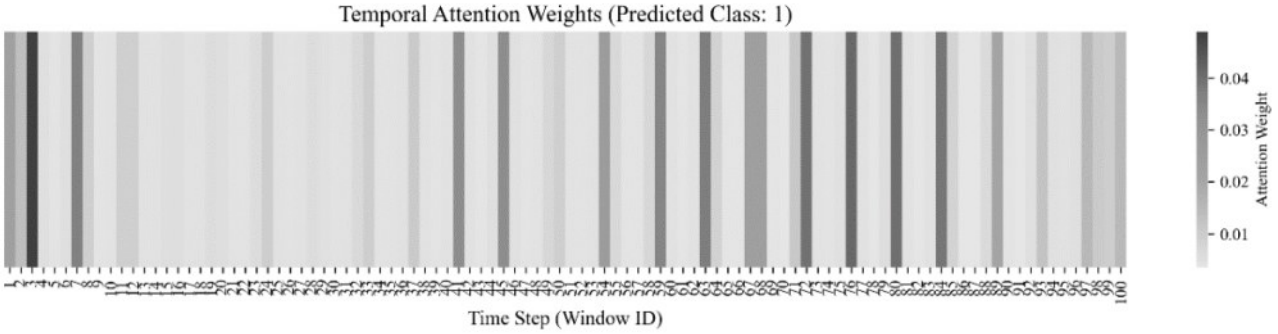



图8　时间维度热力图

#### 4.4.1　注意力机制的语义归因

模型通过集成的 Attention 层，能够自动识别并强化流量序列中对判别“恶意”起关键作用的时间步。如图 8 所示，横轴代表观测期间的 100 个连续滑动窗口，颜色深度反映了模型分配的权重比例

实验验观察发现，权重的分布呈现出显著的“脉冲式”特征，即在浅色背景(代表常规心跳或背景噪声)中，夹杂着若干深红色竖线(如 Time Step 3, 7, 41, 59 等)。这种不均匀分布契合了恶意软件的“突发性”本质：恶意 APP 通常在长时间潜伏后，突然在极短的时间窗口内执行高频 API 调用或敏感数据回传等高危操作。通过这种机制，模型不仅能给出判决结果，还能精准定位恶意行为露出的“马脚”，显著提升了检测的透明度。

#### 4.4.2　全局特征重要性评估

为量化多尺度特征在整体分类决策中的贡献度，本文引入基于协同博弈论的 SHAP 解释框架，SHAP 通过计算各特征分量的沙普利值，量化其相对于基准预测值的边际贡献。如图 9 所示，展示了四类基于 OWASP 的威胁语义特征($p_{c_k}$)对模型输出的边际贡献。

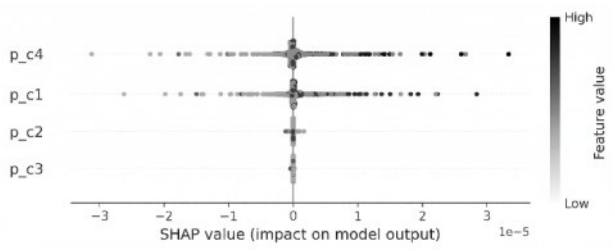



图9　全局语义类型分布特征重要性评估图

图中显示，$p_{c_1}$和$p_{c_4}$在全局样本中拥有最广泛的 SHAP 值分布。当这两个特征的取值较高（图中深色数据点）时，SHAP 值显著为正，强烈驱动模型输出“恶意”判决。这一逻辑高度符合现代移动恶意软件的攻击特征：恶意 APP 为规避静态查杀，往往采用高度混淆或非标准化的 API 通信结构，同时其核心变现手段（如银行木马、窃密软件）高度依赖于对用户敏感凭证的劫持与窃取。

图中$p_{c_1}$和$p_{c_4}$的蓝色数据点（代表该类特征出现频率低），部分浅色点依然产生了正向的 SHAP 值。这表明模型并没有退化为简单的“关键词匹配器”。当恶意软件试图通过降低敏感 API 调用频率来实施“低慢速”潜伏时，模型能够结合其他维度的特征，依然做出准确的风险归因。这种从底层流量统计到高层威胁语义的映射，证明了模型并非盲目拟合数据分布，而是掌握了一定的恶意行为的内在演化规律。

#### 4.4.3　局部决策溯源分析

为了深入剖析模型对单一高危告警样本的判定逻辑，本文针对特定恶意样本执行了局部特征归因分析。

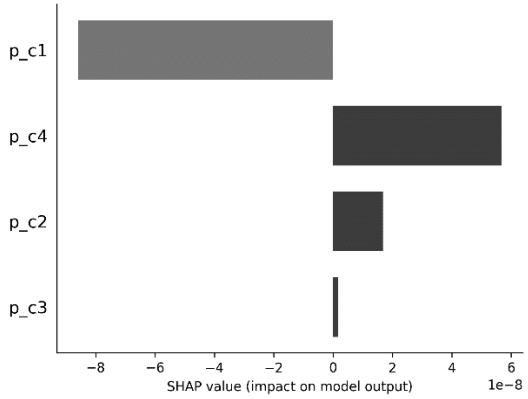



图10　特定恶意样本局部归因分析图

观察该判定窗口内的特征博弈过程，可以发现$p_{c_1}$特征呈现出显著的负向 SHAP 值。根据前文建立的威胁语义映射体系$p_{c_1}$主要对应伪造接口或拦截凭证传输等敏感行为。该特征的负向贡献提示，该样本在当前阶段对本地敏感权限和凭证接口的调用处于较低水平。结合移动恶意软件的常见运行规律推测，这可能是样本为规避启发式检测或沙箱审查而采取的延迟执行或潜伏策略。然而，这种表象上的低危特征并未主导最终的分类决策。同时，$p_{c_4}$与$p_{c_2}$特征产生了较高的正向 SHAP 值，构成了触发恶意告警的主要驱动因素。其中，$p_{c_4}$的正向激增表明模型在流量中捕获到了大量基于路径-参数模板细化的异常结构请求，这往往与恶意软件维持隐蔽命令与控制通信的网络行为相契合；而$p_{c_2}$的正向权重则进一步提示，该应用在网络层面上表现出异常加载第三方资源或注入未知代码的倾向。上述“隐蔽结构化通信”结合“外部动态加载”的特征分布，与移动恶意软件中常见的动态载荷下发机制的底层逻辑基本一致。上述局部溯源分析表明，检测模型并非单纯依赖单一的高危 API 调用频率进

行判决，而是能够在多维特征的相互博弈中，从网络流量的深层结构关联中提取具有实际安全业务意义的异常模式。

### 4.5　核心组件消融实验

为避免复杂网络结构的盲目堆叠，本文采用控制变量法设计了消融实验，验证VMD平稳化与Attention注意力机制的具体性能增益，四组衍生模型均在相同超参数下进行测试。,

实验评估指标计算公式如下

准确率(Accuracy)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

精确率(Precision)

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

召回率(Recall)

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

F1值(F1-score)为精确率与召回率的调和平均数，综合反映模型对恶意类别的检测效果，计算公式为

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

实验数据表明，仅基于基础BiLSTM的模型的加权F1-Score仅为0.8529。在引入Attention机制后，模型F1-Score提升至0.8853，证明动态聚焦策略有效降低了长序列中高危行为的漏报率。在引入VMD模块进行频域解耦后，模型F1-Score显著跃升至0.9226，充分证明了VMD能够有效剥离非平稳网络流量中的无序高频抖动，为时序建模提供纯净输入。本文最终提出的完整模型取得了最优表现，准确率、精确率、召回率与F1-Score分别达到0.9818、0.9376、0.9454与0.9415，证实了“VMD频域解耦降噪”与“Attention时域特征聚焦”具有高度互补的协同效应。

### 4.6　综合对比分析

为进一步客观评价本模型的综合竞争力，将其与现有主流检测框架在CICMalDroid2020数据集上进行横向对比。

实验结果显示：相比于达到99.89%准确率但极度依赖完整动态沙箱环境提取470维特征的AntDroidNet模型，本文方法专注于非侵入式的网络流元数据提取，工程实时性更强。相比于准确率为97.38%的静态图像化分析模型BlockDroid，本文的自适应多尺度窗口有效弥补了静态特征在时间跨度感知上的不足，能够识别跨越长周期的潜伏趋势。此外，面对传统的CNN-LSTM(准确率94.00%)与计算开销巨大的BERT-GPT2大模型(准确率90.75%)，本文方法在VMD与Attention的加持下取得了98.18%的准确率与94.15%的F1值，在保持高识别精度的同时，实现了精度与推理时延的平衡。

## 5　结束语

本文提出一种融合多尺度语义与VMD-BiLSTM的检测模型。模型通过自适应滑动窗口与VMD平稳化解耦优化特征输入，结合Attention-BiLSTM与焦点损失实现隐蔽特征的聚焦，并引入SHAP框架以提供决策溯源。实验表明，该模型在CICMalDroid2020数据集上的准确率与加权F1-Score分别达到98.18%和94.15%，有效兼顾了恶意流量识别的高效性与可解释性。

**表2**　　**消融实验结果(基于CICMalDroid2020)**

| 包含模块 | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| 仅BiLSTM | 0.8950 | 0.8610 | 0.8450 | 0.8529 |
| Attention-BiLSTM | 0.9210 | 0.8920 | 0.8787 | 0.8853 |
| VMD-BiLSTM | 0.9530 | 0.9250 | 0.9202 | 0.9226 |
| **VMD-Attention-BiLSTM (本文)** | **0.9818** | **0.9376** | **0.9454** | **0.9415** |

**表 3　　不同模型效果对比**

| 模型 | 特征类型 | 指标结果 |
|---|---|---|
| AntDroidNet[18] | 动态行为特征 (ACO 优化) | Accuracy=99.89% FPR=0.13% |
| BlockDroid[19] | 静态特征 (DEX 文件图像化) | Accuracy=97.38% F1=96.09% |
| CNN-LSTM[20] | 网络流量统计特征 | Accuracy=94.00% FPR=3.00% |
| BERT-GPT2[21] | 行为语义特征(Transformer) | Accuracy=90.75% F1=91.00% |
| **本文方法** | **APP 网络流量特征** | **Accuracy=98.18% F1=94.15%** |

## 参考文献：


[1] BRYNIELSSON J, SHARMA R. Detectability of low-rate HTTP server DoS attacks using spectral analysis[C]//Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Piscataway: IEEE Press, 2015: 954-961.

[2] CHENG M, LI Q, LV J M, et al. Multi-scale LSTM model for BGP anomaly classification[J]. IEEE Transactions on Services Computing, 2021, 14(3): 765-778.

[3] WANG J Y, WANG Z, LI J F, et al. Multilevel wavelet decomposition network for interpretable time series analysis[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 2437-2446.

[4] FOULADI R F, ERMIŞ O, ANARIM E. A DDoS attack detection and countermeasure scheme based on DWT and auto-encoder neural network for SDN[J]. Computer Networks, 2022, 214: 109140.

[5] ALBAHAR M A. Recurrent neural network model based on a new regularization technique for real-time intrusion detection in SDN environments[J]. Security and Communication Networks, 2019, 2019: 1-9.

[6] PEI J M, ZHONG K Y, JAN M A, et al. Personalized federated learning framework for network traffic anomaly detection[J]. Computer Networks, 2022, 209: 108906.

[7] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]//Proceedings of International Conference on Learning Representations. Vancouver: ICLR Press, 2018: 1-19.

[8] GEIGER A, LIU D Y, ALNEGHEIMISH S, et al. TadGAN: time series anomaly detection using generative adversarial networks[C]// Proceedings of IEEE International Conference on Big Data (Big Data). Piscataway: IEEE Press, 2020: 33-43.

[9] PATIL R, BIRADAR R, RAVI V, et al. Network traffic anomaly detection using PCA and BiGAN[J]. Internet Technology Letters, 2022, 5(1): e235.

[10] 邹福泰, 谭越, 王林, 等. 基于生成对抗网络的僵尸网络检测[J]. 通信学报, 2021, 42(7): 95-106.

[11] CHEN X H, DENG L W, HUANG F T, et al. DAEMON: unsupervised anomaly detection and interpretation for multivariate time series[C]// Proceedings of IEEE 37th International Conference on Data Engineering. Piscataway: IEEE Press, 2021: 2225-2230.

[12] 麻文刚, 张亚东, 郭进. 基于 LSTM 与改进残差网络优化的异常流量检测方法[J]. 通信学报, 2021, 42(5): 23-40.

[13] CHOUHAN N, KHAN A, KHAN H U R. Network anomaly detection using channel boosted and residual learning based deep convolutional neural network[J]. Applied Soft Computing, 2019, 83: 105612.

[14] YANG S. Anomaly traffic detection based on LSTM[C]//Proceedings of IEEE 10th Joint International Information Technology and Artificial Intelligence Conference. Piscataway: IEEE Press, 2022: 667-670.

[15] ULLAH I, MAHMOUD Q H. Design and development of RNN anomaly detection model for IoT networks[J]. IEEE Access, 2022, 10: 62722-62750

[16] WANG S, YAN Q, CHEN Z, et al. TextDroid: Semantics-based detection of mobile malware using network flows[C]//2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2017: 18-23.

[17] ZAMAN M, SIDDIQUI T, AMIN R, et al, 2015. Malware detection in Android by network traffic analysis[C]//Proceedings of 2015 International Conference on Networking Systems and Security, NSysS 2015: 1-5.

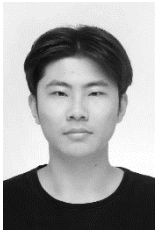
[18] AL OGAILI R R N, RAHEEM O A, ABDKHALEQ M H G, et al. AntDroidNet cybersecurity model: A hybrid integration of ant colony optimization and deep neural networks for android malware detection[J]. Mesopotamian Journal of CyberSecurity, 2025, 5(1): 104-120.

[19] ŞAFAK E, DOĞRU İ A, BARIŞÇI N, et al. BlockDroid: detection of Android malware from images using lightweight convolutional neural network models with ensemble learning and blockchain for mobile devices[J]. PeerJ Computer Science, 2025, 11: e2918.

[20] ANSORI D B, SLAMET J, GHUFRON M Z, et al. Android malware classification using gain ratio and ensembled machine learning[J]. International Journal of Safety and Security Engineering, 2024, 14(1): 259-266.

[21] DJÈ BI DJÈ G G, DIAKO D J, KANGA K, et al. Innovation in cyber threat detection: transformer-based approach[J]. International Journal of Advanced Research, 2024, 12(11): 1375-1389.

**许国良**(1973-)，男，江苏南京人，博士，重庆邮电大学教授，主要研究方向为计算机视觉、大数据分析与挖掘等。



**时磊**(2000-)，男，河南郑州人，主要研究方向为机器学习。

**邱思琦**(1997-)，男，湖北孝感人，主要研究方向为机器学习。

**许宇**(2003-)，男，四川成都人，主要研究方向为机器学习.。