LSTM-GAN时空特征融合的DDoS攻击早期预测方法

Early Prediction of DDoS Attacks Based on Spatiotemporal Feature Fusion with LSTM-GAN

杨 飞[1]，周 晗[2]，由志远[1]，王 新[1]（1. 中讯邮电咨询设计院有限公司，北京 100048；2. 安徽理工大学，安徽 合肥 230041）
Yang Fei[1], Zhou Han[2], You Zhiyuan[1], Wang Xin[1]（1. China Information Technology Designing & Consulting Institute Co., Ltd., Beijing 100048, China; 2. Anhui University of Science and Technology, Hefei 230041, China）

摘 要：
针对DDoS分布式拒绝服务攻击早期检测的时效性与准确性，提出一种基于时空特征融合的LSTM-GAN混合预测模型。通过构建双通道特征提取模块同步捕获网络流量数据的时间和空间关联特征，实现攻击特征的跨维度融合。在对抗训练框架下，通过引入GAN生成对抗网络机制，借助生成器模拟攻击流量演变模式，驱动判别器提升对攻击初期流量变异系数小于5%、持续时间不足10 s的微小波动特征的敏感性。该方法可在攻击流量未形成显著峰值时实现早期预警，为主动式网络安全防护提供新的技术路径。

关键词：
DDoS攻击检测；GAN；LSTM；流量行为分析；GAT；对抗训练
doi：10.12045/j.issn.1007-3043.2025.09.003
文章编号：1007-3043（2025）09-0014-06
中图分类号：TN915.08
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
It addresses the timeliness and accuracy requirements for early detection of DDoS attacks by proposing an LSTM-GAN hybrid prediction model based on spatiotemporal feature fusion. A dual-channel feature extraction module is constructed to synchronously capture temporal and spatial correlation characteristics in network traffic data, achieving cross-dimensional fusion of attack features. Under an adversarial training framework, the introduction of a GAN mechanism enables the generator to simulate the evolution patterns of attack traffic, thereby driving the the discriminator to enhance its sensitivity to the tiny fluctuation features of initial attack traffic with a coefficient of variation <5% and a duration < 10 seconds. This method facilitates early warning when attack traffic has not yet formed significant peaks, providing a novel technical approach for proactive cybersecurity defense. The proposed methodology offers a new pathway for implementing preventive network security protection strategies.

Keywords：
DDoS attack detection; GAN; LSTM; Network traffic behavior analysis; GAT; Adversarial training

引用格式：杨飞，周晗，由志远，等. LSTM-GAN时空特征融合的DDoS攻击早期预测方法[J]. 邮电设计技术，2025(9)：14-19.

1 概述

DDoS分布式拒绝服务攻击是网络安全领域最具破坏性的威胁之一。传统基于阈值统计或签名匹配的检测方法难以应对日益复杂的新型攻击模式，尤其在攻击早期阶段，微弱的异常信号往往淹没在海量正常流量中，制约了防御系统的及时响应能力[1]。

本文提出一种基于时空特征融合的LSTM-GAN混合模型。构建双通道特征提取网络：在时间维度，采用多尺度LSTM捕获流量序列的周期规律；在空间维度，设计GAT图注意力机制解析节点拓扑关联特征[2]。通过对抗训练策略，生成器合成具有时空一致性的攻击演化样本，判别器则结合真实流量进行特征验证，实现对早期攻击特征的放大识别[3]。

收稿日期：2025-07-25

2 系统整体架构设计

本文所提模型由双通道特征提取网络与动态对抗训练框架构成（见图1），其核心目标是通过时空特征联合建模与对抗样本生成，实现对DDoS攻击早期特征的放大识别。

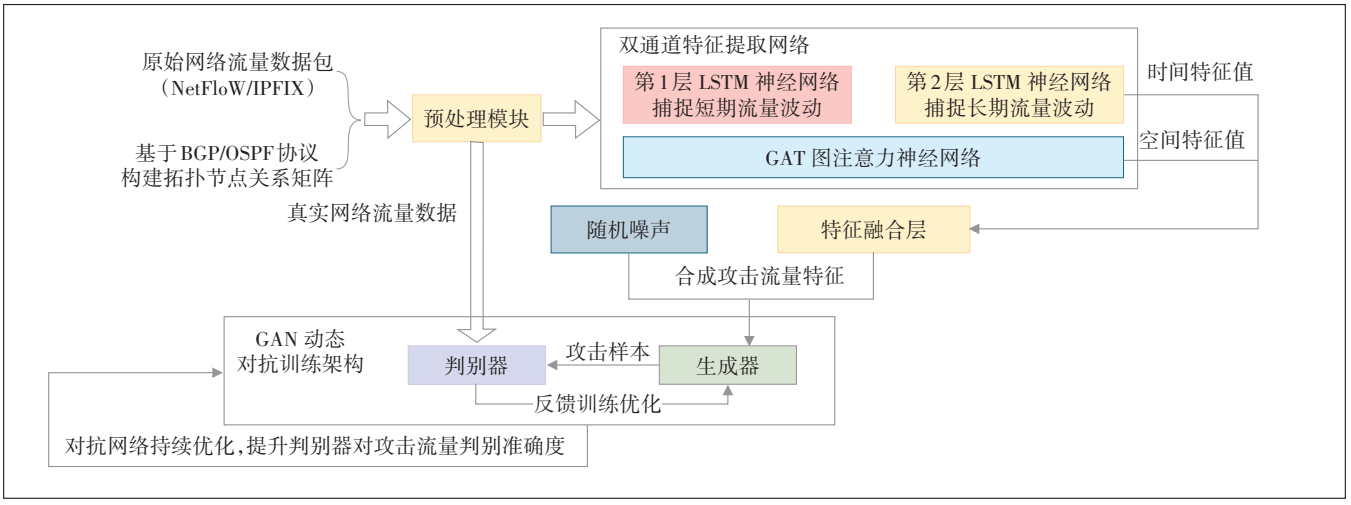

图1　时空特征融合的LSTM-GAN混合模型架构

其中，输入数据经过预处理后，分别进入LSTM神经网络通道和GAT图注意力机制网络，提取时间特征和空间特征；之后，通过特征融合层得到流量的时空特征数据。生成器基于融合特征数据和随机噪声合成攻击样本；判别器则同时验证攻击样本的时序与拓扑合理性，最终通过动态对抗训练优化模型[4]。以下分别介绍各模块功能。

2.1 输入层与预处理模块

预处理模块从NetFlow/IPFIX数据包中提取时间序列特征，并得到原始时间序列数据，之后采用5 s滑动窗口对连续的时间序列数据进行流量序列分隔，每个窗口包含500~1 000条NetFlow/IPFIX记录，为下一步进入LSTM神经网络做好数据预处理准备。

预处理模块从BGP/OSPF协议原始数据中提取空间特征，分别是节点数据和边关系数据。每个节点通过IP地址或AS编号分配唯一标识符；边关系数据构建节点间的连接关系。边的权重根据链路带宽、延迟、跳数或路由策略优先级进行设定[4]。为下一步空间特征数据进入GAT神经网络做好准备。

2.2 双通道特征提取网络

双通道特征提取网络是整个模型的核心，由时间、空间特征通道并行构成挖掘攻击行为的关键特征。

2.2.1 LSTM时间特征通道

时间特征通道设计的目的是为了捕捉网络流量的时序变化规律。该特征通道采用LSTM神经网络作为提取时间维度特征值的模型设计。

LSTM长短期记忆网络是一种时间循环神经网络，是为了解决一般的RNN循环神经网络存在的长期依赖问题而专门设计出来的。一般的RNN循环神经网络存在梯度爆炸和梯度消失的问题，对于长距离的数据的学习效果不好。而LSTM长短记忆网络将信息存储在一个个记忆细胞中，不同隐藏层的记忆细胞之间通过少量线性交互形成一条传送带。同时引入一种“门”的结构，用来新增或删除记忆细胞中的信息，控制信息的流动（见图2）。

传统基于LSTM网络的方法可能忽略瞬时异常[5-6]。本文提出的双层LSTM+注意力机制能识别短时脉冲（见图3），关联长周期攻击阶段，提升对早期攻击信号的敏感度[7]。

第1层（短时特征）：处理秒粒度的流量数据，用于分析诸如SYN Flood的脉冲式攻击短至数秒内的突发流量。每个LSTM单元接收当前时刻的流量数据，并传递隐藏状态以记忆前几秒的流量状态。短时特征LSTM单元计算如下：

$$h_t^{(1)} = \mathrm{LSTM}(X_t,\ h_{t-1}^{(1)}),\ t \in [1,\ 60] \tag{1}$$

第2层（长时特征）：对第1层的输出进行降采样，每10 s取一个特征，捕捉类似HTTP慢速攻击这种缓慢增长的攻击。长时特征LSTM单元计算如下：

$$h_t^{(2)} = \mathrm{LSTM}(h_t^{(1)},\ h_{t-10}^{(2)}) \tag{2}$$

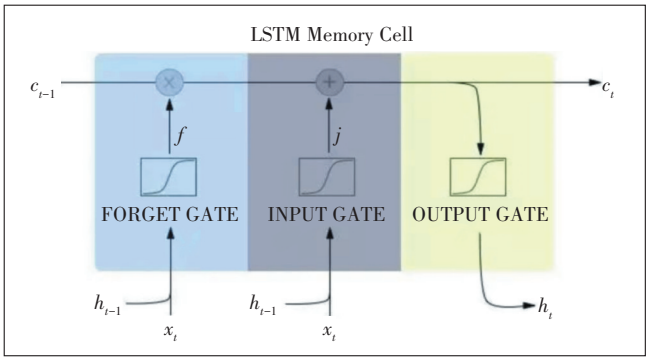

图2 LSTM长短期记忆网络结构图(a)

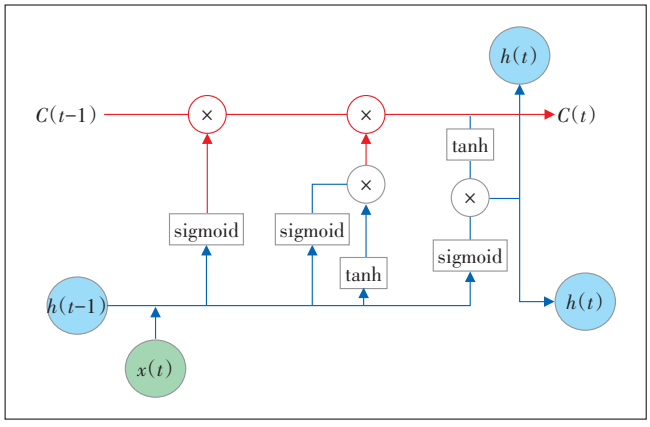

图3 LSTM长短期记忆网络结构图(b)

时间梯度注意力机制:在LSTM模型的每个时间步中,使用一个全连接层来计算注意力权重(见图4)。本文通过注意力机制放大关键突变点的权重,避免攻击初期真正的攻击流量数据被淹没的情况[7]。流量变化率的显著性权重计算公式如下:

$$\alpha_t = \text{soft}\max\left(V^T \tanh\left(W\left(h_t^{(2)} \times \frac{\partial X_t}{\partial t}\right)\right)\right) \quad (3)$$

2.2.2 GAT空间特征通道

空间特征通道专注于挖掘网络拓扑中节点间的关联模式与异常传播路径,通过动态图建模与注意力机制实现攻击行为的空间定位[8]。本文通过GAT图注意力机制网络实现。

首先,GAT是在GNN图神经网络上增加注意力机制形成的神经网络,而GNN本身是一种基于图结构的深度学习方法[9]。如图5所示,GNN图神经网络包含3个核心要素:节点表示,即将每个节点映射到一个低维向量空间中;图结构表示,表示整个图的拓扑结构;消息传递,即节点与其邻居交换信息来更新自身的表示。

本文在GNN图神经网络基础上,增加使用注意力机制,构建GAT图注意力网络(见图6)。GAT的核心工作原理是通过注意力机制来计算节点间的关系。每个节点的状态更新会考虑到其邻居节点的状态,GAT会计算一个节点与其邻居节点之间的边权重,计算公式如下:

$$W_{ij}^{(t)} = \frac{\text{Count}(V_i \rightarrow V_j)}{T} + \alpha \times \text{Topo\_Sim}(V_i, V_j) \quad (4)$$

其中,$\text{Topo_Sim}(v_i, v_j)$表示基于BGP/OSPF路由跳数的相似度,跳数越少,值越大;T表示时间窗口长度,默认值为60,单位为s;α表示拓扑先验权重系数,默认值为0.3。

GAT由堆叠在一起的图注意力层构成,每个图注

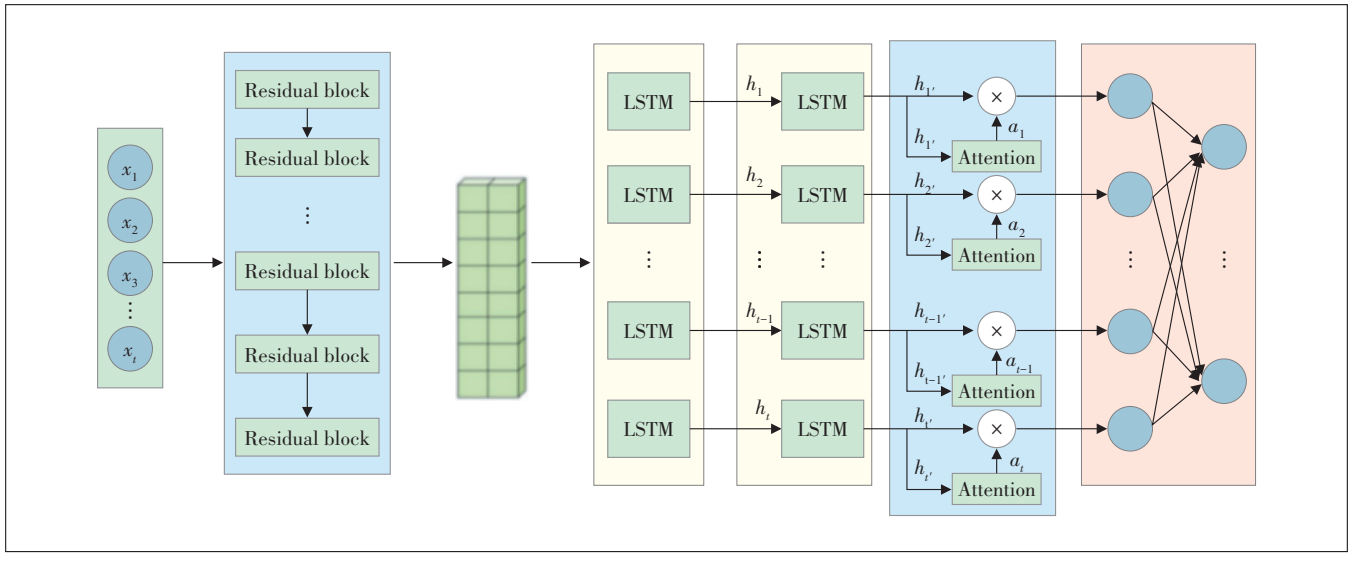

图4 LSTM时间特征通道

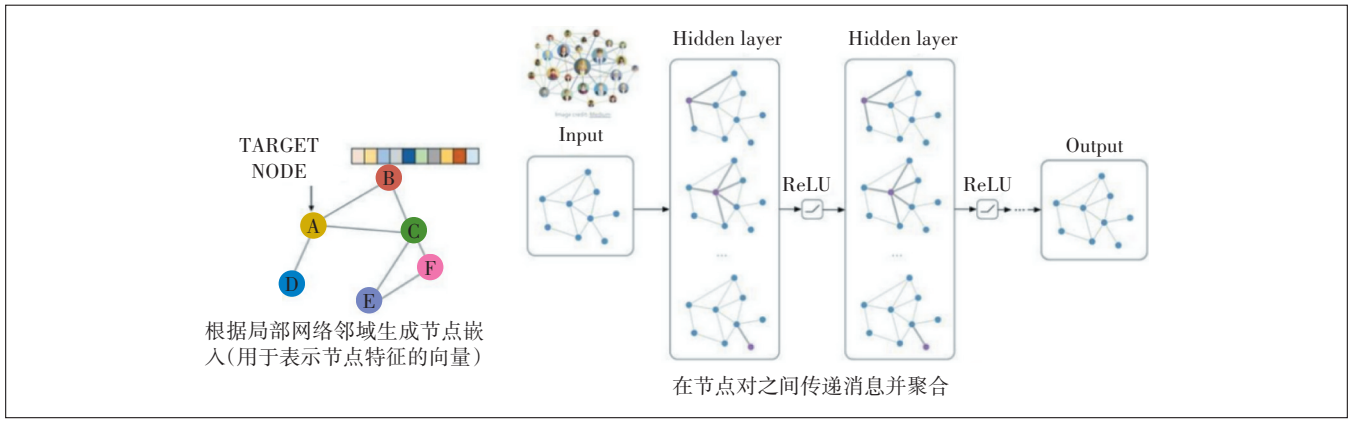

图5　GNN图神经网络

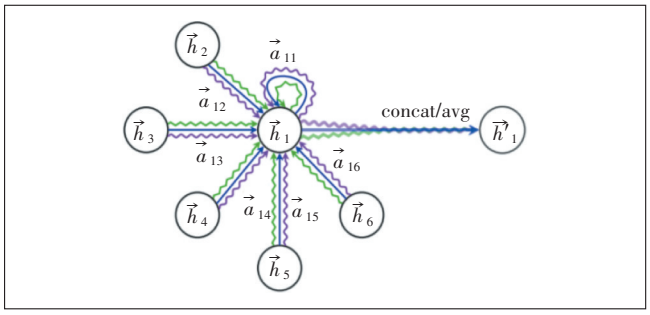

图6　GAT图注意力机制网络

意力层获取节点嵌入作为输入。通过邻居信息聚合，使每个节点通过多头注意力机制，从其直接相连的邻居节点中筛选重要信息[9]，注意力机制 h_i' 的计算公式如下：

$$h_i' = V_{i,i}Wh_i + \sum_{j \in N(i)} V_{i,j}Wh_j \tag{5}$$

其中，$V_{i,j}$ 的计算公式如下：

$$V_{i,j} = \text{LeakyReLU}(\alpha^T[Wh_i \| Wh_j]) \tag{6}$$

若某节点突然与大量新IP通信，如10 s内新增连接IP数超过基线均值3倍以上，其邻居中边权重>0.6的连接会传递异常信号[8]。这样采用多头图注意力机制，关注与当前节点流量突变高度同步的邻居，从而放大异常信号。

2.3 特征融合层

LSTM网络输出的时间特征与GAT网络输出的空间特征，分别通过全连接层调整维度，确保特征长度对齐。然后通过门控融合机制进行动态权重分配，并设计基于Sigmoid函数的可学习门控单元，根据时空特征的语义相关性生成融合权重，计算公式如下[10-11]：

$$Z_{\text{fusion}} = \eta \times Z_{\text{time}} + (1-\eta) \times Z_{\text{space}},\ \eta \in [0,\ 1] \tag{7}$$

其中，Z_{time} 表示提取的时间特征值，Z_{space} 表示提取的空间特征值；η 为全连接层输出门控权重值，计算公式如下：

$$\eta = \sigma(W_g[Z_{\text{time}} + Z_{\text{space}}] + b_g) \tag{8}$$

其中，σ 为Sigmoid函数。

在将时间和空间特征值进行融合处理后，将数据输出到GAN生成对抗网络，完善攻击流量的判别器，下文将重点描述GAN网络的执行过程。

2.4 GAN生成对抗网络

GAN生成对抗网络本质是一种深度敏感词模型，其核心结构由生成器和判别器2个部分构成[12-13]。生成器的核心任务是学习真实数据的分布特征，生成与真实数据高度相似的新样本；判别器则扮演鉴别角色，通过分析输入数据的统计特性，判断其来源于真实数据集还是生成器的合成结果[3]。

在训练过程中，生成器不断优化生成策略以欺骗判别器，使其无法分辨合成数据的真实性；而判别器则持续提升鉴别能力，力求精准区分2类数据源[6]。这种对抗性机制推动两者性能交替提升，最终得到用于判别DDoS攻击流量的判别器。

2.4.1 生成器

本文GAN对抗网络中的生成器核心任务是通过学习真实网络流量数据，生成具有时空关联性的合成数据，辅助判别器更早地识别DDoS攻击[14]。随机噪声向量作为初始输入，该噪声向量代表潜在空间中的随机特征，本文使用高斯噪声向量：$G \sim N(0,\ 1)$。

结合上文提到的时空特征编码在特征融合层处理的结果 Z_{fusion}，最后将噪声向量和特征融合向量合成输入参数，输入到生成器中：

$$G_{\text{in}} = \text{Concat}(Z_{\text{fusion}},\ G) \tag{9}$$

2.4.2 生成器对抗训练优化

生成器在对抗训练中的优化过程是一个动态博弈的闭环系统,本文使用生成器接收判别器的空间异常分数,通过策略梯度的强化学习机制调整攻击源节点策略,形成“生成—判别—反馈”闭环。策略梯度的计算公式如下:

$$\nabla J = G_{\text{in}} \times \sum_{t=0}^{T} \gamma^t (1 - S_t) \tag{10}$$

其中,$r^t = 1 - S_t$,而 S_t 为判别器反馈的异常分数,下文会有介绍。通过这种反馈机制,不断优化生成器生成攻击样本的过程(见图7)。

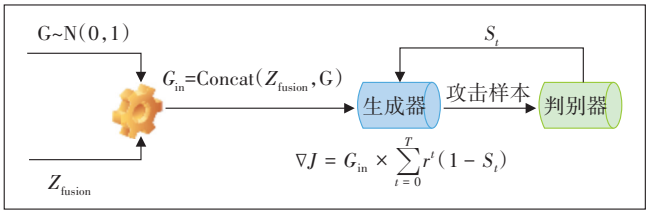

图7 GAN生成器对抗训练优化

2.4.3 判别器

判别器的核心目标是学习区分真实网络流量与生成器合成的样本流量。判断输入数据是来自真实流量还有生成器伪造的样本流量。当判别器判定正确率分布通过训练越来越高时,面对真实网络则更容易判别是否为网络攻击[2,15]。

判别器结构通常包含输入数据、特征深度抽象、对抗决策输出和反馈驱动优化(见图8)。

a)输入数据层。接收原始流量数据,使其分别进入LSTM网络和GAT网络进行特征提取。其中时间特征由LSTM提取的网络流量时序模式;而空间特征则通过GAT网络提取的流量空间关联。

b)特征深度抽象。通过堆叠的全连接层,逐步抽象出输入特征的高阶表示,第1层捕获流量统计量,深层网络整合跨时间窗的流量趋势与多维空间关联,识别复杂攻击模式。

c)对抗决策输出。决策网络使用CNN卷积神经网络,最终通过Sigmoid激活函数,输出一个0~1的标量值,表示输入数据属于真实攻击流量分布的概率。决策公式如下:

$$D(x) = \frac{1}{1 + e^{-(W_d^T h^{(L)} + b_d)}} \tag{11}$$

其中,W_d^T 和 b_d 是CNN卷积决策网络的权重与偏置参数,$h^{(L)}$ 是输入数据层传入的流量数据。经过CNN卷积最后一层Sigmoid运算后,使得 $D(x) \in (0,1)$ 作为判别器的输出。

d)反馈驱动优化。在训练过程中,当误判真实流量为假时,调整参数以提高对真实模式的敏感度;这个过程通过向生成器反馈异常分数 S_t 完成,其计算公式如下:

$$S_t = \alpha \times (x - G_{\text{out}}) + \beta \times \log D(G_{\text{out}}) \tag{12}$$

其中,α 和 β 是超参数,用于平衡重构误差与对抗性损失的贡献,G_{out} 是生成器产生并传入到判别器中的样本参数。

2.5 实验测试

2.5.1 实验数据选择及预处理

本文采用公开网络流量数据集CIC-DDoS2019,实验数据涵盖正常流量与多种DDoS攻击类型,如:TCP-SYN Flood、UDP Flood。同时,实验数据集还覆盖流量统计特征、协议类型、时间戳等关键字段。

将实验数据集划分为训练集、验证集、测试集三大类。首先通过训练集训练模型,并通过验证集对训练过程中的结果进行验证,使模型最终达到稳定状态,最后通过测试集数据对训练好的模型进行实验测试。

2.5.2 早期预警有效性测试

本文从模型结构和攻击发起时间2个维度分别进行预警有效性测试。

在模型结构对比维度上,使用不同模型方法,在测试集上模拟攻击场景。通过数据集中的时间戳确定攻击开始时间,针对不同模型统一使用攻击发起前

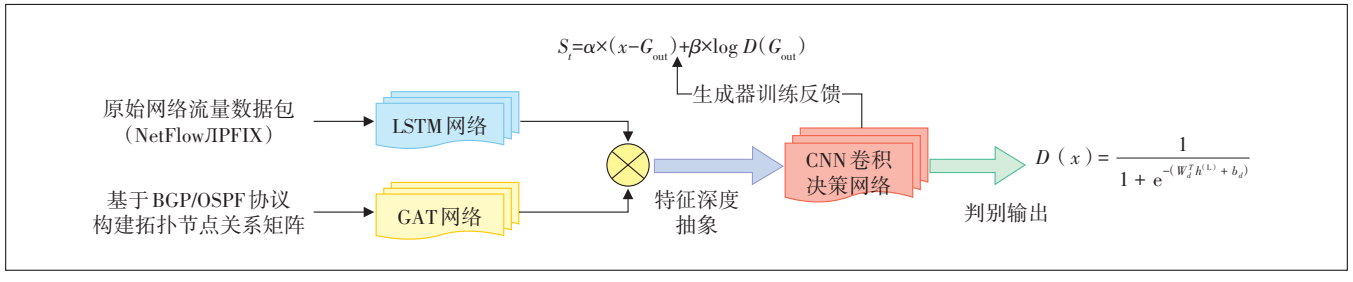

图8 GAN判别器结构

10 min为基准测试，统计所有流量中，被正确识别为攻击流量的比例，即准确率；统计正常流量中被误判为攻击的比例，即误报率；统计攻击流量没有识别报警的比例，即漏报率；记录模型从攻击发生到触发警报的时间差，即检测延迟。测试结果如表1所示。

表1　早期预警有效性(a)

模型名称	准确率/%	误报率/%	漏报率/%	平均检测延迟/s
LSTM-GAN(本文模型)	98.5	0.8	1.2	3.2
纯LSTM模型	92.8	2.7	2.4	7.1
CNN-LSTM时空融合模型	94.7	1.7	1.9	5.8
传统阈值检测方法	89.1	3.5	5.2	13.5

在攻击发起时间维度上，通过区分攻击发起前10 min、5 min、1 min分别进行测试，单独测试本文LSTM-GAN模型的早期预警能力，结果如表2所示。

表2　早期预警有效性(b)

预测时间窗口/min	准确率/%	误报率/%	漏报率/%	平均检测延迟/s
10	98.5	0.8	1.2	3.2
5	97.8	1.1	1.5	2.8
1	96.2	1.6	2.3	1.5

从表2可以看出，本文模型在攻击前10 min即达到98.5%的准确率，表明时空特征融合可有效捕捉攻击前10 min内流量时序变异系数>1.2或节点拓扑连接熵增加0.5以上的早期异常。随着预测时间窗口缩短，检测延迟降低至1.5 s，但误报率略有上升，说明该模型需要权衡实时性与准确性，以达到最佳预测状态。

2.5.3　时空特征消融测试

为验证时空特征双通道在模型中的作用，采用时空通道开关的方式分别进行测试，即使用以下3种场景分别进行测试：时空特征双通道的LSTM-GAN模型、仅使用空间特征通道设置LSTM网络权重值矩阵$W_{LSTM}=[0]$、仅使用时间特征通道设置GAT图注意力机制网络权重值矩阵$W_{GAT}=[0]$。测试使用攻击发起前10 min的数据为基准进行，结果如表3所示。

从表3可以看出，时空特征融合在相同攻击时间基准的条件下，可以有效提升整个模型的预测准确性，降低误报率和漏报率。当仅时间特征时，准确率下降6%，误报率上升3.4%；当仅空间特征时，准确率下降7.8%，漏报率上升6.3%；当时空融合时，F1-score提升8.2%。可见无论减少时间维度的特征提取，还是减少空间维度的特征提取，都会降低整体模型的性能指标。因此采用时空双维度融合十分必要。

表3　时空特征消融实验

特征组合	检测准确率/%	误报率/%	漏报率/%
完整时空特征+LSTM-GAN	98.5	0.8	1.2
仅时间特征(LSTM)	92.5	4.2	6.3
仅空间特征(GAT)	90.7	5.8	7.5

参考文献：

[1] 李志强，王芳. 面向DDoS攻击的时空特征融合与动态阈值预警模型[J]. 信息网络安全，2025，25(4)：33-42.
[2] 周涛，徐静. 时空图卷积与LSTM融合的DDoS攻击早期预警模型[J]. 计算机研究与发展，2025，62(6)：1345-1356.
[3] 王磊，刘伟. BiTCN-Transformer与LSTM-GAN的并行预测模型[J]. 自动化学报，2025，51(12)：2785-2796.
[4] 周凯，刘杰. 基于时空张量分解的网络攻击态势预测方法[J]. 计算机应用，2025，45(12)：3721-3730.
[5] 赵婵，张瑞生. 轻量级LSTM-GAN模型在边缘计算场景下的DDoS检测研究[J]. 计算机研究与发展，2025，62(8)：1789-1800.
[6] 赵婵，张瑞生. 基于RNN-LSTM-GAN混合模型的网络流量异常预测[J]. 计算机学报，2025，48(3)：567-578.
[7] 吴敏，黄浩. 基于注意力机制的LSTM-GAN网络流量预测优化[J]. 计算机工程与应用，2025，61(9)：156-165.
[8] 陈思，杨帆. 时空特征与流量基线的联合建模在DDoS检测中的应用[J]. 通信技术，2025，58(5)：102-110.
[9] 陈刚，周洋. 深度学习驱动的网络流量时空特征分解方法[J]. 自动化学报，2025，51(8)：1789-1800.
[10] 刘洋，高杰. 多变量时间序列异常检测中的GAN-LSTM混合模型研究[J]. 网络与信息安全学报，2025，11(3)：78-89.
[11] 陈刚，周洋. 基于混合损失函数的LSTM-GAN模型训练优化[J]. 计算机工程与应用，2025，61(12)：156-165.
[12] 张华，吴磊. 基于流量熵值波动与LSTM-GAN的DDoS攻击检测[J]. 计算机科学，2025，52(12)：301-310.
[13] 周涛，徐静. 端到端时空特征学习在网络安全预测中的创新应用[J]. 计算机研究与发展，2025，62(10)：2105-2116.
[14] 李明，王伟. 基于时空特征融合的LSTM-GAN网络流量预测框架[J]. 软件学报，2025，36(4)：1023-1035.
[15] 李志强，王芳. 动态阈值与时空特征联合建模的DDoS早期预警[J]. 信息网络安全，2025，25(5)：78-87.

作者简介：
杨飞，毕业于合肥学院，高级工程师，学士，主要从事网络安全技术的研究工作；周晗，毕业于中国科学技术大学，副教授，硕士，主要从事网络安全技术的研究和教学工作；由志远，毕业于西安电子科技大学，高级工程师，硕士，主要从事网络软件研发工作；王新，毕业于北京航空航天大学，硕士，主要从事网络安全技术方向的研究工作。