融合溯源图与知识图谱的 APT 攻击检测模型研究

安　渊[1]，鲍永庆[2]

（1. 国家计算机网络应急技术处理协调中心西藏分中心，西藏　拉萨　850000；
2. 中共西藏自治区委员会网络安全和信息化委员会办公室，西藏　拉萨　850000）

摘　要：针对高级持续性威胁（APT）攻击所具有的隐蔽性强、持续时间长、多阶段渐进的特点，提出了一种融合动态系统行为溯源图与静态威胁情报知识图谱的检测模型。该模型使用时空图注意力网络联合建模攻击链中的空间依赖与时间演化关系。通过图注意力网络捕捉实体间可疑关联，通过门控循环单元建模行为序列的阶段性演进，从而实现对 APT 攻击全链条的端到端检测。在 Windows-APTs Dataset 2025 公开数据集上的实验表明，所提模型在 APT 多分类检测任务中性能良好，准确率达 95. 14%，F1 分数为 95. 29%。

关键词：APT 攻击检测；溯源图；知识图谱

中图分类号：TP393. 08　　**文献标志码：**A　　**DOI：**10. 19358/j. issn. 2097 - 1788. 2026. 03. 002

中文引用格式：安渊，鲍永庆. 融合溯源图与知识图谱的 APT 攻击检测模型研究［J］. 网络安全与数据治理，2026，45(3)：10 - 16.

英文引用格式：An Yuan，Bao Yongqing. Research on an APT attack detection model integrating provenance graphs and knowledge graphs［J］. Cyber Security and Data Governance，2026，45(3)：10 - 16.

Research on an APT attack detection model integrating provenance graphs and knowledge graphs

An Yuan[1]，Bao Yongqing[2]

（1. National Computer Network Emergency Response Technical Team/Coordination Center of China，Xizang Branch，Lhasa 850000，China；
2. Office of the Cyberspace Administration and Informatization Committee of the Communist Party of China Xizang Autonomous Region Committee，Lhasa 850000，China）

Abstract：Advanced Persistent Threat（APT）attacks，characterized by strong concealment，long duration，and multistage progressive patterns，were addressed by a novel detection model. The model was constructed through the fusion of dynamic system behavior provenance graphs with static threat intelligence knowledge graphs. Spatial dependencies and temporal evolution relationships within attack chains were jointly modeled using spatial-temporal graph attention networks. Suspicious associations between entities were captured through graph attention mechanisms，while stage-wise evolution of behavioral sequences was modeled using gated recurrent units，enabling end-to-end detection of complete APT attack chains. Experiments on the public Windows-APTs Dataset 2025 demonstrated that the proposed model performed well in the APT multi-classification detection task，with an accuracy of 95. 14% and an F1-score of 95. 29%.

Key words：APT attack detection；provenance graph；knowledge graph

0　引言

高级持续性威胁（Advanced Persistent Threat，APT）攻击因其隐蔽性、持续性和组织化特征，已经成为企业级网络安全的核心挑战。区别于传统的网络攻击，APT 攻击通常由具备明确战略意图的组织发起，采用多阶段、渐进式的攻击模式，综合运用社会工程学、零日漏洞利用及复杂的命令与控制网络，旨在长期潜伏并窃取高价值信息[1]。传统依赖已知特征码匹配或基于单点异常阈值的检测方法[2]，因其缺乏对攻击全局上下文和内在逻辑关联的理解，往往难以奏效，导致漏报与误报。

为突破这一瓶颈，基于系统审计日志构建数据溯源图[3]的研究范式应运而生。该方法通过将分散的系统事件重构为具有因果与时间属性的有向图，能够直观地刻画攻击链中实体间的依赖关系，为还原复杂的多步攻击提供了强大的结构化表示基础。

与此同时，知识图谱技术为整合与利用网络安全领域的碎片化信息提供了理想框架。特别是以 MITRE ATT&CK[4]为代表的知识库，系统化地建模了 APT 组织、攻击技术、利用工具及防御措施之间的复杂关联。

1 研究现状

1.1 现有工作

数据溯源图通过将系统实体抽象为节点，将实体间的交互行为抽象为有向边，并保留精确的时间戳，为重现攻击链提供了天然的表示形式。

早期的 APT 检测方法多依赖专家规则和签名匹配[5]，虽可解释性较高，但难以应对未知攻击与复杂多变的 APT 行为。随着数据规模与攻击隐蔽性的提升，现有的研究重心逐步转向基于特征工程与统计学习的方法[6]，通过提取网络流量、日志序列或行为统计特征，结合支持向量机、决策树、随机森林等模型进行分类。这类方法虽然提升了一定的自动化水平，但仍受限于特征表达能力，难以捕捉 APT 攻击中跨阶段、长周期的复杂依赖关系。

近年来，深度学习与序列建模方法成为主流，其核心在于重点关注 APT 攻击的时序性与行为累积效应[7-8]。文献［9－11］中，利用长短期记忆网络（Long Short-Term Memory，LSTM）、双向长短期记忆网络（Bidirectional Long Short-Term Memory，BiLSTM）等建模网络流量或溯源图序列，通过长短记忆提取攻击的长期特征。并且文献［11］中进一步引入注意力机制增强关键行为捕捉能力，有效提升了长周期隐蔽攻击的识别效果。

随着图数据建模能力的突破，基于图结构与知识图谱的方法也成为主流方法之一。这类方法将系统实体、网络流量、威胁情报等抽象为图结构，利用图卷积网络（Graph Convolutional Network，GCN）、图同构网络（Graph Isomorphism Network，GIN）、深度卷积图神经网络（Deep Convolutional Graph Neural Network，DCGNN）等进行高阶关系推理[9,12-17]。文献［17］提出了 RAS-GNN 模型，针对 APT 攻击中警报缺失和碎片化的问题，利用图注意力机制重构攻击场景，通过补全攻击链条中的隐式关联来提升检测的鲁棒性。文献［9］提出 BiADG 模型，将 BiLSTM 与 DCGNN 结合，兼顾序列特征与图结构关系。文献［14－16］分别从异构信息网络推理、攻击路径图追踪、多视图图表示等角度，提升了检测的可解释性与场景覆盖能力。

此外，针对 APT 数据不平衡、攻击场景多样等挑战，文献［13］中使用生成对抗网络（Generative Adversarial Networks，GAN）生成对抗样本，文献［10－11，18－20］整合网络、日志、DNS 等多维数据，分别从数据增强、多源信息融合、异常检测与正常行为建模[11,21]等角度进行探索。

现有的研究工作呈现出从规则驱动到数据驱动、从单点检测到全链条分析、从特征工程到表示学习的发展趋势。然而，在多源异构数据的深度融合与统一表示方面仍有局限。现有图模型在动态、异构图结构建模与实时推理效率方面仍然存在挑战。同时，对 APT 攻击的跨阶段、多步骤行为进行端到端、可解释的关联分析能力仍显不足。这些挑战为本文基于动态多视图异构图网络、支持多源信息融合的 APT 检测框架提供了研究动机与创新空间。

本文提出一种融合溯源图与知识图谱的 APT 攻击检测模型（Multi-G-Sentry），通过动态－静态双图融合机制，实现了系统溯源图与先验威胁情报知识图谱在特征层面的深度融合。在此基础上，构建了一个统一的时空图注意力网络（Spatio-Temporal Graph Attention Network，ST-GAT）模型。利用图注意力网络[22]（Graph Attention Network，GAT）层在融合后的单帧图谱中学习实体间可疑依赖的空间权重，并通过门控循环单元[23]（Gated Recurrent Unit，GRU）层跨时间维度建模攻击行为的阶段性演进，从而实现对 APT 攻击链空间依赖性与时间延续性的联合捕获与精准分析。

1.2 技术路线对比分析

本文对比了近年来具有代表性的三类 APT 检测模型，包括：基于时间序列的 TSE-APT[7]、基于单一溯源图的 RAS-GNN[17]、基于异构图融合的 HGNN-CTI[18]，以及所提 Multi-G-Sentry 模型，如表 1 所示。

表 1 不同 APT 检测模型的特性对比

模型	核心架构	时空建模机制	知识融合
TSE-APT	Ensemble LSTM	仅时间	无
RAS-GNN	Graph Attention	仅空间	无
HGNN-CTI	Heterogeneous GNN	弱时序	简单拼接
Multi-G-Sentry	GAT + GRU	联合时空	特征对齐

现有的主流检测方法主要的局限在于时空特征割裂，以及外部知识利用不足。本文所提 Multi-G-Sentry 模型引入 CTI 知识图谱作为静态先验，通过 GAT 层自适应地聚合威胁情报语义。

基于 APT 检测所存在的难题，本文主要的研究工作如下：

（1）设计了一个端到端的 Multi-G-Sentry 模型，将

动态的系统行为溯源图与静态的网络安全知识图谱进行特征级的深度融合。该范式使模型兼具基于异常行为的感知能力与基于威胁情报的验证能力。

（2）设计了一种结合 GAT 与 GRU 的时空图神经网络架构。GAT 层赋予模型对图中关键恶意关联的聚焦能力，GRU 层使模型能够记忆并关联跨时间窗口的攻击步骤，从而有效刻画长周期攻击。

（3）通过消融实验定量验证了 Multi-G-Sentry 模型在准确率、精确率、召回率以及 F1 分数上的显著提升。

2　系统模型

2.1　动态时空图网络下的 APT 攻击检测问题定义

APT 攻击检测本质上是一个复杂的时空图分类任务。给定一段时间内的系统审计日志流 $L=(l_1, l_2, \cdots, l_T)$，通过构建溯源图的方式将其转换为一系列时间切片图 $\boldsymbol{G}=(\boldsymbol{G}_1, \boldsymbol{G}_2, \cdots, \boldsymbol{G}_T)$，其中每个图 $\boldsymbol{G}_t$，$t=1, 2, \cdots, T$ 表示一个时间窗口内的所有实体间关系，T 为时间窗口数量。为了增强模型对已知攻击模式的识别能力，将威胁情报知识图谱（CTI-KG）融入到每个节点的特征表示中，使得每个图节点都包含丰富的背景信息和潜在风险评估。模型需要学习映射函数 f：$\boldsymbol{G}\rightarrow y$，其中 $y\in\{0, 1, \cdots, C-1\}$ 表示 APT 攻击的类别标签，C 为总类别数。

Multi-G-Sentry 模型采用三层架构设计，包括数据输入层、ST-GAT 模块和分类决策层。

2.2　数据输入层

在 Multi-G-Sentry 框架中，数据输入层主要包括两部分：动态溯源图序列和 CTI-KG。

2.2.1　动态溯源图序列

使用公开数据集 Windows-APTs Dataset 2025（https：//data. mendeley. com/datasets/b8fmtzvpy8/2），通过解析脚本将原始审计日志转换为标准的 CSV 格式，每行记录了一个系统活动事件，包含的字段有：_source（源实体），_target（目标实体），_relation（实体间的关系类型），_timestamp（事件发生的时间戳）。将所有事件按照时间窗口进行分组，时间窗口大小为 Δt，分组后转化为一系列有向图 $\boldsymbol{G}=(\boldsymbol{G}_1, \boldsymbol{G}_2, \cdots, \boldsymbol{G}_T)$。

2.2.2　CTI-KG

从 MITRE ATT&CK 框架获取威胁情报知识图谱，用于增强节点特征。该知识图谱包含 APT 组织、攻击技术、恶意工具以及相互关系的信息。CTI-KG 作为外部知识库，通过实体对齐和嵌入学习，为动态溯源图中的节点提供语义丰富的特征表示。

2.3　ST-GAT 时空特征提取与融合

如图 1 所示，Multi-G-Sentry 采用端到端的层次化架构，主要包含数据输入、GAT、GRU 以及融合分类层。ST-GAT 模型主要由空间与时间两个特征提取阶段组成。输入的特征图首先由 GAT 层处理，学习节点间的空间注意力权重；其输出序列进而由 GRU 层处理，建模攻击行为的阶段性演进。两个分支的特征在融合层进行拼接，作为下游分类的依据。

2.3.1　空间特征提取：GAT

对于每个时间窗口内的溯源图 $\boldsymbol{G}_t=(\boldsymbol{V}_t, \boldsymbol{E}_t)$，其中 $\boldsymbol{V}_t$ 表示节点集合，$\boldsymbol{E}_t$ 表示边集合。给定节点特征矩阵 $\boldsymbol{X}_t\in\mathbb{R}^{|V_t|\times d}$，其中 d 为融合后的特征维度。GAT 通过多头注意力机制学习节点空间依赖关系。

对于图中任意相邻节点对 $(i, j)\in\boldsymbol{V}_t$，计算原始注意力系数：

$$e_{ij}=\mathrm{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}_{\mathrm{space}}\boldsymbol{h}_i\|\boldsymbol{W}_{\mathrm{space}}\boldsymbol{h}_j]) \tag{1}$$

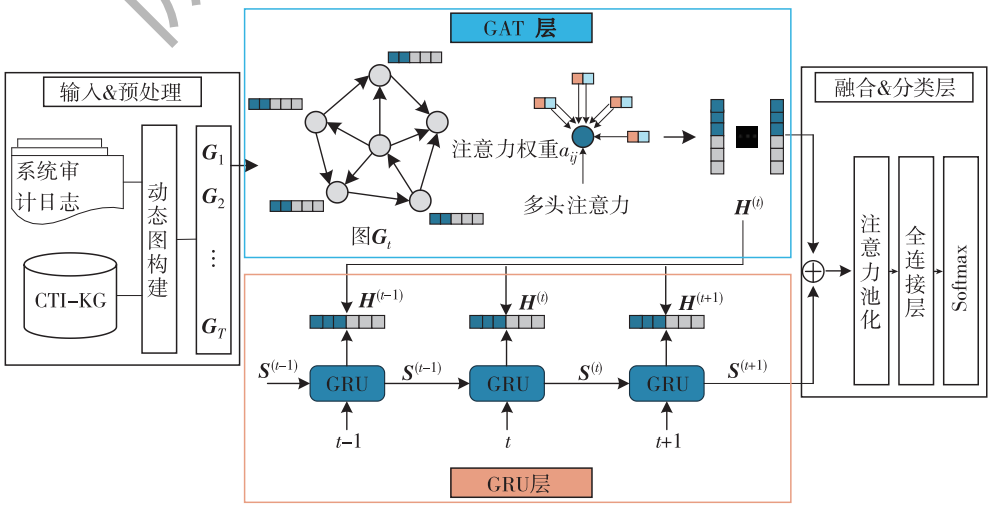

图 1　时空图注意力网络 ST-GAT

其中 LeakyReLU(·) 为带泄露的 ReLU 激活函数，$\boldsymbol{a} \in \mathbb{R}^{2d}$为注意力向量，$\boldsymbol{W}_{\text{space}} \in \mathbb{R}^{d \times d}$为节点空间特征的线性变换矩阵，$\boldsymbol{h}_i$，$\boldsymbol{h}_j \in \mathbb{R}^{d}$ 为节点 i 和节点 j 的输入特征向量，$\|$表示向量拼接。为了使注意力系数具有可比性，需对节点 i 的所有邻居 $j \in \boldsymbol{N}(i)$ 的注意力系数进行归一化处理：

$$a_{ij} = \text{Softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{l \in \boldsymbol{N}(i)} \exp(e_{il})} \tag{2}$$

其中 a_{ij}表示节点 i 对邻居 j 的“关注程度”，$\boldsymbol{N}(i)$ 表示节点 i 的邻居节点集合。节点 i 的空间特征输出为：

$$\boldsymbol{h}_i^{\text{space}} = \sigma\left(\sum_{j \in \boldsymbol{N}(i)} a_{ij} \boldsymbol{W}_{\text{space}} \boldsymbol{h}_j\right) \tag{3}$$

其中 $\sigma(\cdot)$ 为非线性激活函数。

考虑到 APT 攻击的空间关联往往具有多维度的特性，单头注意力机制可能无法全面捕捉节点间的复杂关系。因此，GAT 引入多头注意力机制，通过并行计算多个独立的单头注意力，融合多维度的空间关联信息。假设使用 K 个注意力头，计算注意力头 k，$k \in \{1, 2, \cdots, K\}$ 的注意力系数 e_{ij}^k和归一化系数 a_{ij}^k，以及节点 i 在注意力头 k 的空间特征输出 $\boldsymbol{h}_i^{\text{space},k}$：

$$e_{ij}^k = \text{LeakyReLU}\left((\boldsymbol{a}^k)^{\text{T}} \left[\boldsymbol{W}_{\text{space}}^k \boldsymbol{h}_i^{\text{space}} \| \boldsymbol{W}_{\text{space}}^k \boldsymbol{h}_j^{\text{space}}\right]\right) \tag{4}$$

其中 $\boldsymbol{a}^k \in \mathbb{R}^{2d_{\text{space}}}$为注意力头 k 的注意力向量；$\boldsymbol{W}_{\text{space}}^k \in \mathbb{R}^{d_{\text{space}} \times d_{\text{space}}}$为注意力头 k 的线性变换矩阵，$\boldsymbol{h}_i^{\text{space}}$，$\boldsymbol{h}_j^{\text{space}} \in \mathbb{R}^{d_{\text{space}}}$为节点 i 和节点 j 的空间特征向量；d_{space}为融合后的空间特征维度。同样，注意力头 k 下，对节点 i 的所有邻居 $j \in \boldsymbol{N}(i)$ 的注意力系数进行归一化处理：

$$a_{ij}^k = \text{Softmax}_j(e_{ij}^k) \tag{5}$$

节点 i 在注意力头 k 下的空间特征输出为：

$$\boldsymbol{h}_i^{\text{space},k} = \sigma\left(\sum_{j \in \boldsymbol{N}(i)} a_{ij}^k \boldsymbol{W}_{\text{space}}^k \boldsymbol{h}_j^{\text{space}}\right) \tag{6}$$

将 K 个注意力头的输出拼接得到最终的空间特征：

$$\boldsymbol{h}_i^{\text{space}} = \|_{k=1}^{K} \boldsymbol{h}_i^{\text{space},k} \tag{7}$$

此时，$\boldsymbol{h}_i^{\text{space}} \in \mathbb{R}^{K \cdot d_{\text{space}}}$表示节点 i 与邻居的多维度空间依赖关系。

在完成空间特征提取后，ST-GAT 需进一步捕捉溯源图序列中的时间演化模式。APT 攻击通常呈现多阶段、长周期的特性，仅依赖空间关联无法识别此类复杂行为，因此，ST-GAT 采用 GRU 对节点空间特征序列进行时序建模。

2.3.2 时间特征提取：GRU

给定时间窗口 t 内节点 i 的空间特征序列 $\boldsymbol{H}_i = \{\boldsymbol{h}_{i,1}^{\text{space}}, \boldsymbol{h}_{i,2}^{\text{space}}, \cdots, \boldsymbol{h}_{i,T}^{\text{space}}\}$，其中 $\boldsymbol{h}_{i,t}^{\text{space}} \in \mathbb{R}^{d_{\text{space}}}$。GRU 通过更新门 $\boldsymbol{z}_{i,t}$和重置门 $\boldsymbol{r}_{i,t}$控制信息传递。更新门 $\boldsymbol{z}_{i,t} \in [0, 1]$ 决定保留多少历史信息，同时也决定融入多少当前空间特征：

$$\boldsymbol{z}_{i,t} = \text{Sigmoid}\left(\boldsymbol{W}_z \boldsymbol{h}_{i,t}^{\text{space}} + \boldsymbol{U}_z \boldsymbol{h}_{i,t-1}^{\text{hidden}} + \boldsymbol{b}_z\right) \tag{8}$$

其中 $\boldsymbol{W}_z \in \mathbb{R}^{d_{\text{hidden}} \times K \cdot d_{\text{space}}}$为更新门的输入权重矩阵，$\boldsymbol{U}_z \in \mathbb{R}^{d_{\text{hidden}} \times d_{\text{hidden}}}$为更新门的隐藏状态权重矩阵，$\boldsymbol{h}_{i,t-1}^{\text{hidden}} \in \mathbb{R}^{d_{\text{hidden}}}$为上一时间窗口的隐藏状态，$\boldsymbol{b}_z \in \mathbb{R}^{d_{\text{hidden}}}$为偏置项，$d_{\text{hidden}}$为时间特征维度。重置门 $\boldsymbol{r}_{i,t} \in [0, 1]$ 决定忽略多少历史信息，同时也决定关注多少当前空间特征：

$$\boldsymbol{r}_{i,t} = \text{Sigmoid}\left(\boldsymbol{W}_r \boldsymbol{h}_{i,t}^{\text{space}} + \boldsymbol{U}_r \boldsymbol{h}_{i,t-1}^{\text{hidden}} + \boldsymbol{b}_r\right) \tag{9}$$

其中 $\boldsymbol{W}_r \in \mathbb{R}^{d_{\text{hidden}} \times K \cdot d_{\text{space}}}$为重置门的输入权重矩阵，$\boldsymbol{U}_r \in \mathbb{R}^{d_{\text{hidden}} \times d_{\text{hidden}}}$为重置门的隐藏状态权重矩阵，$\boldsymbol{b}_r \in \mathbb{R}^{d_{\text{hidden}}}$为偏置项。基于重置门的输出，计算当前时间窗口的候选隐藏状态：

$$\tilde{\boldsymbol{h}}_{i,t} = \tanh\left(\boldsymbol{W}_h \boldsymbol{h}_{i,t}^{\text{space}} + \boldsymbol{U}_h \left(\boldsymbol{r}_{i,t} \otimes \boldsymbol{h}_{i,t-1}^{\text{hidden}}\right) + \boldsymbol{b}_h\right) \tag{10}$$

其中 $\tanh(\cdot)$为双曲正切激活函数，$\boldsymbol{W}_h \in \mathbb{R}^{d_{\text{hidden}} \times K \cdot d_{\text{space}}}$为候选状态的权重矩阵，$\boldsymbol{b}_h \in \mathbb{R}^{d_{\text{hidden}}}$为偏置项，$\otimes$为逐元素乘法，$\boldsymbol{r}_{i,t} \otimes \boldsymbol{h}_{i,t-1}^{\text{hidden}}$表示对历史信息进行筛选，仅保留与当前时间窗口相关的部分。那么，当前时间窗口的隐藏状态 $\boldsymbol{h}_{i,t}^{\text{hidden}}$通过更新门融合历史信息和候选隐藏状态：

$$\boldsymbol{h}_{i,t}^{\text{hidden}} = (1 - \boldsymbol{z}_{i,t}) \otimes \boldsymbol{h}_{i,t-1}^{\text{hidden}} + \boldsymbol{z}_{i,t} \otimes \tilde{\boldsymbol{h}}_{i,t} \tag{11}$$

2.3.3 时空特征融合

时间特征提取已捕获节点活动的时间演化规律，空间特征提取已量化节点间的空间依赖关系。为全面捕捉 APT 攻击的时空特性，需将节点级的时空特征融合为图级表示，以实现对整个系统状态的攻击判断。对于每个时间窗口 t 内的节点 i，其节点级时空特征由空间特征与时间特征采用特征拼接策略融合：

$$\boldsymbol{h}_{i,t}^{\text{spatio-temp}} = \left[\boldsymbol{h}_{i,t}^{\text{space}} \| \boldsymbol{h}_{i,t}^{\text{hidden}}\right] \tag{12}$$

其中 $\boldsymbol{h}_{i,t}^{\text{spatio-temp}} \in \mathbb{R}^{K \cdot d_{\text{space}} + d_{\text{hidden}}}$为节点 i 在时间窗口 t 下的时空特征。进一步，将节点级时空特征聚合为图级表示。采用注意力机制对节点特征进行加权求和，自适应突出与攻击更为相关的节点。计算节点 i 的注意力权重：

$$a_{i,t} = \text{Softmax}_i\left(\boldsymbol{\omega}^{\text{T}} \cdot \tanh\left(\boldsymbol{W}_{\text{att}} \cdot \boldsymbol{h}_{i,t}^{\text{spatio-temp}} + \boldsymbol{b}_{\text{att}}\right)\right) \tag{13}$$

其中 $\boldsymbol{W}_{\text{att}} \in \mathbb{R}^{d_{\text{att}} \times (K \cdot d_{\text{space}} + d_{\text{hidden}})}$为注意力层的线性变换矩阵，$d_{\text{att}}$为注意力机制的隐藏维度，$\boldsymbol{\omega} \in \mathbb{R}^{d_{\text{att}}}$为注意力向量，$\boldsymbol{b}_{\text{att}} \in \mathbb{R}^{d_{\text{att}}}$为偏置项。基于注意力权重，图 $\boldsymbol{G}_t$ 的图级时空特征的加权求和为：

$$\boldsymbol{h}_{G_t}^{\text{graph}} = \sum_{i \in V_t} a_{i,t} \cdot \boldsymbol{h}_{i,t}^{\text{spatio-temp}} \tag{14}$$

其中 $\boldsymbol{h}_{G_t}^{\text{graph}} \in \mathbb{R}^{K \cdot d_{\text{space}} + d_{\text{hidden}}}$ 融合了图 $\boldsymbol{G}_t$ 的全局时空状态。

2.4 分类决策层

将图级时空特征输入全连接分类层，输出当前时间窗口对应 APT 攻击的类别概率分布。分类层首先将图级特征映射到 C 维空间：

$$\boldsymbol{s}_t = \boldsymbol{W}_c \cdot \boldsymbol{h}_{G_t}^{\text{graph}} + \boldsymbol{b}_c \tag{15}$$

其中 $\boldsymbol{s}_t \in \mathbb{R}^C$，$\boldsymbol{W}_c \in \mathbb{R}^{C \times (K \cdot d_{\text{space}} + d_{\text{hidden}})}$ 为分类层权重矩阵，$\boldsymbol{b}_c \in \mathbb{R}^C$ 为分类层偏置项。将 $\boldsymbol{s}_t$ 映射为概率分布，得到每个类别的预测概率：

$$\hat{\boldsymbol{y}}_t = \text{Softmax}(\boldsymbol{s}_t) \tag{16}$$

其中 $\hat{\boldsymbol{y}}_t \in \mathbb{R}^C$，且满足 $\sum_{c=0}^{C-1} \hat{\boldsymbol{y}}_t[c] = 1$。

3 实验结果分析

为验证 Multi-G-Sentry 在 APT 多分类检测任务中的有效性，基于公开数据集 Windows-APTs Dataset 2025 进行试验。该数据集包含 36 个模拟中国威胁行为者的攻击场景。其攻击行为完全与 MITRE ATT&CK 框架对齐，每个模拟攻击步骤都映射到了 ATT&CK 矩阵中的具体战术和技术。

3.1 数据集预处理

对数据进行清洗和特征工程处理，从 Windows-APTs Dataset 2025 的 100 + 个字段中选择 12 个核心特征，包含进程、文件、用户、网络四大实体，以及行为类型、时间戳、威胁情报三大辅助维度，确保特征与 APT 攻击的“实体 – 行为 – 威胁”链条高度相关，如表 2 所示。

表 2 关键特征及字段

特征名称	数据集对应字段名
进程 GUID	_source. data. win. eventdata. processGuid
父进程 GUID	_source. data. win. eventdata. parentProcessGuid
进程命令行	_source. data. win. eventdata. commandLine
目标文件名	_source. data. win. eventdata. targetFilename
文件哈希	_source. data. win. eventdata. hashes
主体用户 SID	_source. data. win. eventdata. subjectUserSid
源 IP	_source. data. win. eventdata. sourceIp
目标 IP	_source. data. win. eventdata. destinationIp
事件 ID	_source. data. win. system. eventID
事件时间戳	_source. @ timestamp
MITRE 战术	_source. rule. mitre. tactic
MITRE 技术	_source. rule. mitre. technique

对数据进行清洗，移除时间戳超出数据集范围的日志、eventID 为空或不在 Windows 安全事件列表的日志，以及 processGuid 或 parentProcessGuid 为空的日志，删除 processGuid + @ timestamp + eventID 完全一致的重复记录，清洗后剩余有效日志 102 011 条。

时间序列构建将处理后的连续事件流转换为适用于动态溯源图模型的输入格式。采用滑动窗口法，基于 APT 攻击阶段的时间特性，选择时间窗口大小 $\Delta t = 30$ min，滑动步长 $s = 10$ min。此步骤将 102 011 条日志记录转换为 8 782 个时间窗口样本，每个窗口对应一个溯源图快照。动态溯源图构建规则如下：

节点：窗口内进程（processGuid）、文件（targetFilename）、用户（subjectUserSid）、网络 IP（destinationIp）；

边：窗口内的行为关联，如“processGuid = P1→targetFilename = F1”表示“进程 P1 访问文件 F1”，“P1→destinationIp = IP1”表示“进程 P1 连接 IP1”；

属性：节点属性包含 commandLine、hashes、mitre. technique，边属性包含 eventID、@ timestamp。

3.2 实验设置

3.2.1 数据划分

经过预处理后的数据集被划分为训练集、验证集和测试集。数据集中 70% 的数据用于模型训练，15% 的数据用于验证，15% 的数据用于测试。

3.2.2 实验环境

本次实验基于 Windows 64 位操作系统，编程语言为 Python 3. 11，深度学习框架为 PyTorch 2. 3. 0，图神经网络使用 Torch Geometric 2. 7. 0。

3.2.3 评价指标

采用准确率（Accuracy）、精确率（Precision）、召回率（Recall）和 F1 分数（F1-score）四个常用的评价指标衡量模型的性能。

3.3 实验结果分析

3.3.1 APT 攻击检测性能分析

经过训练，Multi-G-Sentry 模型在相应的测试集上表现出良好的性能。

如图 2 所示，随着训练轮次的增加，训练准确率与验证准确率基本重合，最终上升至 0. 951 4。训练损失与验证损失基本重合，最终下降至 0. 093 9。

模型在准确率、精确率、召回率及 F1 分数上的表现如表 3 所示，可以看出，Multi-G-Sentry 模型在多维度评价指标下展现出优异性能。

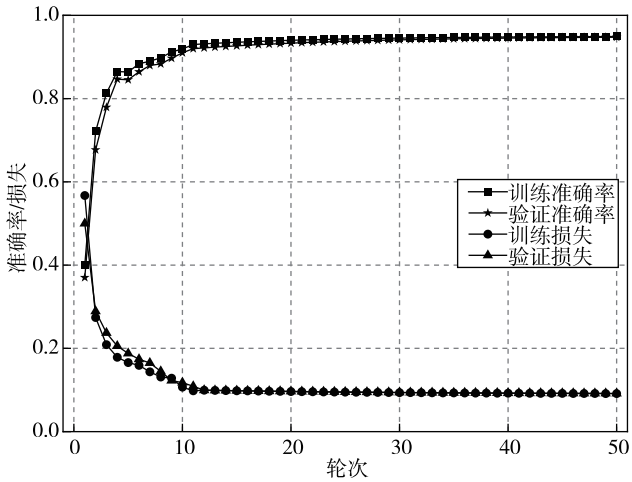

图 2　模型训练性能

表 3　模型整体性能

评价指标	分值	描述
准确率	0.951 4	正确预测的样本数占总样本数的比例
精确率	0.958 2	预测为正类的样本，实际也为正类的比例
召回率	0.947 7	实际为正类的样本，被成功预测为正类的比例
F1 分数	0.952 9	精确率和召回率的调和平均数

3.3.2　消融实验

为了系统评估知识图谱、GAT 以及 GRU 组件对模型性能的贡献，进行消融实验。实验设计与性能表现如表 4 所示。同时为了更直观地展示各组件对模型性能的影响，将移除各组件后的性能指标绘制为柱状图，如图 3 所示。

表 4　消融实验性能对比

模型架构	准确率	精确率	召回率	F1 分数
移除知识图谱	0.879 4	0.885 2	0.871 5	0.878 3
移除 GAT	0.831 3	0.836 5	0.824 0	0.830 2
移除 GRU	0.893 7	0.901 5	0.889 2	0.895 3
Multi-G-Sentry	0.951 4	0.958 2	0.947 7	0.952 9

从消融实验的性能结果可以看到，当移除知识图谱后，模型失去对已知攻击模式的识别能力，识别准确率仅约 88%。当移除 GAT 时，模型无法捕捉节点间的恶意连接，导致误报和漏报大幅增加，识别准确率仅约 83%。移除 GRU 后，模型难以关联跨时序的攻击行为，识别准确率仅约 89%。而 Multi-G-Sentry 模型融合先验知识、空间依赖与时间演化，表现出约 95% 的识别准确率，同时也验证了每个组件在 Multi-G-Sentry 中的必要性。

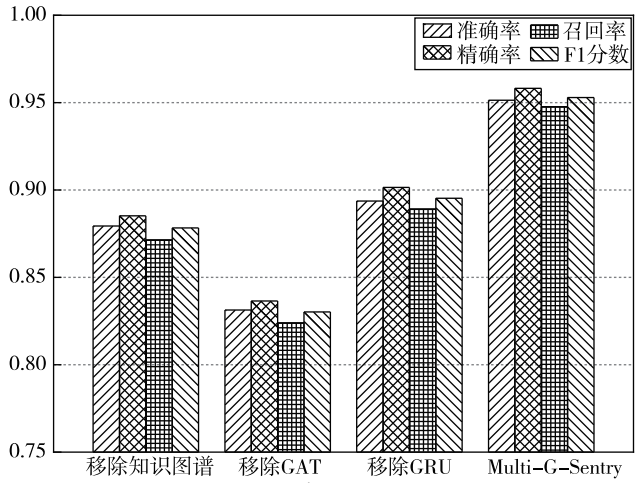

图 3　各组件对性能指标的影响柱状图

4　结束语

本文针对 APT 攻击检测中的隐蔽性与长周期难题，提出了 Multi-G-Sentry 模型。通过构建 ST-GAT 架构，模型有效地融合了动态溯源图的空间拓扑与静态知识图谱的先验语义，并利用 GRU 实现了对攻击阶段演进的精准捕捉。实验结果表明，该模型在 Windows-APTs Dataset 2025 数据集上的 F1 分数达到 95.29%。

尽管模型表现优异，但仍存在一定的局限性。首先，模型的检测性能在一定程度上依赖于 CTI-KG 的完备性。若知识库未能及时更新最新的 APT 组织特征，模型对零日漏洞攻击的关联能力可能会下降。其次，随着监控时间的增长，溯源图规模呈指数级膨胀，基于 GAT 的全图推理在计算资源消耗上较大，在处理超大规模企业级日志时可能面临实时性挑战。

因此，未来的研究方向将集中于以下两个方面：

（1）轻量化与实时性优化：研究基于动态图剪枝和图对比学习的技术，剔除冗余的良性系统噪声，降低 ST-GAT 的计算复杂度，以适应边缘侧网关的实时检测需求。

（2）大语言模型（LLM）赋能的可解释性溯源：针对图神经网络“黑盒”决策难以理解的问题，结合最新的“LLM + 安全”研究趋势[8,12,24]，探索利用 LLM 作为图模型的解释器。通过将溯源图中的攻击路径转化为自然语言描述，辅助安全分析快速理解攻击意图，实现从“威胁检测”到“智能溯源”的跨越。

参考文献

[1] ALSHAMRANI A, MYNENI S, CHOWDHARY A, et al. A survey on advanced persistent threats: techniques, solutions, challenges, and research opportunities [J]. IEEE Communications

Surveys & Tutorials, 2019, 21 (2): 1851 - 1877.
[2] GARCIA-TEODORO P, DIAZ-VERDEJO J E, MACIA-FERNANDEZ G, et al. Anomaly-based network intrusion detection: techniques, systems and challenges [J]. Computers & Security, 2009, 28 (1 - 2): 18 - 28.
[3] HOSSAIN M N, MILAJERDI S M, WANG J, et al. SLEUTH: real-time attack scenario reconstruction from COTS audit data [C]// 26th USENIX Security Symposium (USENIX Security 17), 2017: 437 - 454.
[4] GEORGIADOU A, MOUZAKITIS S, ASKOUNIS D. Assessing MITRE ATT&CK risk using a cyber-security culture framework [J]. Sensors, 2021, 21 (9): 3267.
[5] CHE MAT N I, JAMIL N, YUSOFF Y, et al. A systematic literature review on advanced persistent threat behaviors and its detection strategy [J]. Journal of Cybersecurity, 2024, 10 (1): 275 - 279.
[6] 张奕, 杨晓帆. 基于溯源图分析的高级持续威胁检测技术综述 [J]. 网络安全与数据治理, 2025, 44 (8): 1 - 9.
[7] CHENG M, XIANG G, YANG Q, et al. TSE-APT: an APT attack-detection method based on time-series and ensemble-learning models [J]. Electronics, 2025, 14 (15): 2924.
[8] BENABDERRAHMANE S, VALTCHEV P, CHENEY J, et al. APT-LLM: embedding-based anomaly detection of cyber advanced persistent threats using large language models [J]. arXiv preprint arXiv: 2502. 09385, 2025.
[9] CHO D X, NGUYEN T T. A novel approach for APT attack detection based on an advanced computing [J]. Scientific Reports, 2024, 14 (1): 22223.
[10] AKBARZADEH A, ERDODI L, HOUMB S H, et al. Two-stage advanced persistent threat (APT) attack on an IEC 61850 power grid substation [J]. International Journal of Information Security, 2024, 23 (4): 2739 - 2758.
[11] 梁若舟, 高跃, 赵曦滨. 基于序列特征提取的溯源图上 APT 攻击检测方法 [J]. 中国科学: 信息科学, 2022, 52 (8): 1463 - 1480.
[12] SUN D Y, ZHANG J, XU J, et al. From alerts to intelligence: a novel LLM-aided framework for host-based intrusion detection [J]. arXiv preprint arXiv: 2507. 10873, 2025.
[13] BUCHTA R, GKOKTSIS G, HEINE F, et al. Advanced persistent threat attack detection systems: a review of approaches, challenges, and trends [J]. Digital Threats: Research and Practice, 2024, 5 (4): 1 - 37.
[14] ZHU T, YU J, XIONG C, et al. APTSHIELD: a stable, efficient and real-time APT detection system for Linux hosts [J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20 (6): 5247 - 5264.
[15] CHEN S, CHEN G. APT attack and detection technology [C]// 2024 IEEE 6th Advanced Information Management, Communications, Electronic and Automation Control Conference (IMEC). Piscataway: IEEE, 2024: 795 - 801.
[16] CHO D X, NGUYEN H C. A novel approach for APT attack detection based on feature intelligent extraction and representation learning [J]. PLOS ONE, 2024, 19 (6): e0305618.
[17] HUANG Z, WANG P. RAS-GNN: reconstructing APT attack scenario using graph neural network [C]//ICASSP 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2025: 1 - 5.
[18] MEAD A J, ARABO A. APT attribution using heterogeneous graph neural networks with contextual threat intelligence [J]. Electronics, 2025, 14 (23): 4597.
[19] SACHIDANANDA V, PATIL R, SACHDEVA A, et al. APTer: towards the investigation of APT attribution [C]//2023 IEEE Conference on Dependable and Secure Computing (DSC). Piscataway: IEEE, 2023: 1 - 10.
[20] XIANG G, SHI C, ZHANG Y S, et al. An APT event extraction method based on BERT-BiGRU-CRF for APT attack detection [J]. Electronics, 2023, 12 (15): 3349.
[21] AL-AAMRI A S, ABDULGHAFOR R, TURAEV S, et al. Machine learning for APT detection [J]. Sustainability, 2023, 15 (18): 13820.
[22] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks [J]. arXiv preprint arXiv: 1710. 10903, 2018.
[23] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint arXiv: 1406. 1078, 2014.
[24] ALY A, IQBAL S, YOUSSEF A, et al. Megr-apt: a memory-efficient apt hunting system based on attack representation learning [J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 5257 - 5271.

（收稿日期：2025 - 12 - 10）

作者简介：

安渊（1993 - ），男，本科，工程师，主要研究方向：网络安全、数据安全。

鲍永庆（1990 - ），男，本科，工程师，主要研究方向：网络数据治理、网络安全政策。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部