融合知识图谱与多智能体协同的ATT&CK网络威胁情报问答框架研究

郝耀辉，李娜，陈鑫，秦渝栋，常禾雨
（网络空间部队信息工程大学密码工程学院，河南 郑州 450001）

摘要：为解决现有网络威胁情报问答推理过程中可解释性不足、防御策略生成碎片化问题，提出融合知识图谱与多智能体协同的网络威胁情报问答框架。该框架包含ATT&CK知识图谱、多智能体协同、模型推理增强三大功能模块，通过ATT&CK与STIX 2.1数据构建双向可逆的网络威胁知识图谱，采用黑板架构和LangGraph技术设计多类智能体协同，结合SP-CoT、图检索完成模型推理增强，实现攻击溯源研判和防御策略生成。实验结果表明，在AthenaBench基准测试上，该框架显著优于对比模型，论证了所提框架在多跳推理和防御策略生成方面的优越能力。

关键词：攻击意图识别；知识图谱；ATT&CK；大语言模型；智能体；防御策略
中图分类号：TP309
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB260339

Fusion of knowledge graph and multi-agent collaboration for ATT&CK cyber threat intelligence question-answering framework

Hao Yaohui, Li Na, Chen Xin, Qin Yudong, Chang Heyu
School of Cryptographic Engineering, Information Engineering University, Zhengzhou 450001, China

Abstract: To solve the problems of insufficient interpretability and fragmented defense strategy generation in the reasoning process of existing cyber threat intelligence question-answering systems, a cyber threat intelligence question-answering framework integrating a knowledge graph and multi-agent collaboration was proposed. The framework was composed of three major functional modules: the ATT&CK knowledge graph, multi-agent collaboration, and model inference enhancement. A bidirectionally reversible cyber threat knowledge graph was constructed from ATT&CK and STIX 2.1 data. Collaboration among multiple types of agents was implemented using the blackboard architecture and LangGraph technology. Model inference enhancement was achieved by integrating SP-CoT and graph retrieval, whereby attack attribution analysis and defense strategy generation were enabled. Experiments were carried out on the AthenaBench benchmark, and prominent performance improvements over comparison models were obtained. Experimental results demonstrate that the proposed framework delivers superior capabilities in multi-hop reasoning and defense strategy generation.

Key words: attack intent recognition, knowledge graph, ATT&CK, large language model, agent, defense strategy

收稿日期：2026–06–10；**修回日期**：2026–08–12
通信作者：李娜，19069679580@163.com
基金项目：国家社科基金资助项目（No.21BXW057）；河南省科技攻关基金资助项目（No.252102211040）
Foundation Items：The National Social Science Fundation of China (No.21BXW057), Science and Technology Research Project of Henan Province (No.252102211040)

0 引言

近年来，随着高级持续性威胁、零日漏洞利用和勒索软件等网络攻击手段越来越复杂，国家安全、工业生产和社会稳定面临严峻挑战。网络威胁情报（cyber threat intelligence, CTI）通过收集、分析和整合与威胁行为有关的证据信息，为防御策略制定和攻击行为建模提供关键支撑。然而，现有CTI数据大多存在格式非结构化、知识碎片化、依赖人工标注和语义映射不准确的问题，降低了威胁检测和攻击推理的鲁棒性与准确性[1]。

知识图谱可以从大量分散的网络安全数据中提取实体及其关系，用图结构展示概念、属性和语义关联，并通过知识推理挖掘潜在的攻击模式[2]，在CTI领域展现出巨大潜力。MITRE ATT&CK框架使用战术、技术和子技术的分层结构系统来描述攻击行为，已成为业界公认的攻防知识库。研究人员以ATT&CK为辅助，基于CTI报告构建了多种知识图谱，Noel等[3]提出CyGraph框架，利用ATT&CK的关系结构对攻击行为进行建模，但其主要刻画静态链接关系，难以表达攻击步骤间的动态依赖。Xiang等[4]提出了A4CKGE方法，借助知识表示学习执行多步攻击推理，但其推理过程仍不透明。Li等[5]提出基于GNN的威胁情报知识图谱分析方法，通过节点与边的表征传播挖掘隐含的威胁实体关系。Bilot等[6]提出FEAE方法，结合自监督与少样本学习实现攻击分析。然而，上述方法难以直接输出可追溯的推理路径，可解释性受限。

大语言模型（large language model，LLM）凭借其强大的语言理解与推理能力，为CTI知识图谱的自动化构建与推理提供了新路径。在构建层面，Cheng等[7]提出CTINexus框架，利用上下文学习从CTI报告中抽取实体关系三元组，提供了新思路，但其未从本体设计上解决攻防语义的双向可逆性问题；Wang等[8]提出LLM-CAKG方法，通过多阶段提示工程完成攻击知识图谱的自动化构建和聚类优化，提高了构建效率，但所构建图推理可追溯性不足。基于此，Yang等[9]提出了CTI-Thinker系统通过上下文学习和低秩适配（low-rank adaptation, LoRA）微调提取结构化的威胁实体和关系，用向量对齐策略完成实体归一化和知识融合。Nguyen等[10]测评了LLM从CTI报告中识别攻击技术的实际表现，揭示了类别不平衡、过拟合和领域复杂性等核心挑战，并提出通过两阶段流水线来提高识别精度。Lewis等[11]提出检索增强生成（retrieval-augmented generation，RAG）方法，通过引入外部检索机制提高LLM生成内容的事实准确性，缓解了知识滞后的问题，但直接用于端到端攻击检测时仍存在幻觉问题；Houssel等[12]也发现LLM在精确攻击检测上表现不佳，但结合RAG后性能明显提升，但其图检索因边覆盖率不足常导致推理链断裂。总体而言，现有工作在构建与推理层面多为功能模块的简单拼接，缺少从结构化推理需求出发的系统性机制设计，攻击分析和防御策略生成的客观性与可追溯性都还需要进一步提高。

在上述背景下，智能体技术逐渐成为整合LLM与结构化知识、实现攻击检测与防御推理的重要方法。Wen等[13]提出MobiLLM，利用多智能体实现6G网络的闭环威胁缓解，但其防御策略生成的适应性与通用性还需进一步验证；Oh等[14]提出威胁驱动的多智能体框架，通过LangGraph编排多智能体工作流，实现日志证据ATT&CK行为的映射和策略充分性评估；Deng等[15]提出Pentest-GPT框架，利用自交互模块分工完成渗透测试子任务，验证了LLM智能体在安全场景的可行性，但其属于前馈式架构，攻防难以协同；Liu等[16]进一步探索多智能体协同的事件响应机制，提高了响应决策的协同效率，但防御结果不会反向修正攻击路径判定；Wang等[17]提出多智能体Actor-Critic自主网络防御方法，实现了多智能体协同防御，但缺少对攻击的准确分析。这些研究大多只关注攻击检测或者策略验证其中一个方面，缺少攻击分析与防御策略生成的双向闭环协同。

针对上述问题，本文提出了融合知识图谱与多智能体协同的ATT&CK网络威胁情报问答框架（cyber threat intelligence graph-based question answering，CTI-GQA）。该框架深度结合知识图谱结构化推理和LLM语义理解的优势，旨在实现从攻击实体抽取、攻击链推理到防御方案生成的一体化智能问答与攻防分析。本文主要工作如下。

（1）提出了融合ATT&CK与STIX2.1的类型化双向网络威胁知识图谱构建方法，通过关系类型约束和双向可逆边设计，解决传统图谱动态推理能力不足的问题，实现攻击链多跳路径的结构化推理，提高攻击推理的可解释性。

（2）设计了结构化路径思维链，引入类型映射函数和自一致性校验机制，将知识图谱的多跳路径序列化为可验证的文本推理链，突破传统 CoT 缺少结构化约束的限制。

（3）设计了基于 LangGraph 的多智能体协同架构，实现攻击分析与防御策略生成的双向闭环。

（4）搭建了网络威胁情报专属问答框架，并在自建测试集与 AthenaBench 基准上进行实验验证。实验结果表明，本文提出的多智能体框架有效弥补了传统方法的不足，增强了攻防协同分析能力，为攻击溯源研判、自动化防御策略生成提供了高效可行的技术支撑。

1　问答框架设计

1.1　总体架构

本文提出融合知识图谱与多智能体协同的 ATT&CK 网络威胁情报问答框架 CTI-GQA（如图 1 所示），能够接收用户自然语言形式的威胁情报提问，经过内部多模块协同处理，最终输出可溯源、可解释、结构化的高质量问答结果。该框架由 ATT&CK 知识图谱、模型推理增强、多智能体协同三大核心功能模块组成。ATT&CK 知识图谱模块基于 MITRE ATT&CK 框架构建统一的威胁知识图谱本体，经实体抽取、关系对齐以及知识融合后写入 Neo4j 图数据库，实现知识存储与图谱可视化。模型推理增强模块选用经过 LoRA 微调的 Phi 系列大模型[18]，并设计结构化路径思维链（structured path CoT，SP-CoT）提示工程来进行多跳推理，从而避免只靠黑箱参数知识产生的“幻觉”问题[19]。多智能体协同框架模块依托 LangGraph 构建查询理解、攻击分析、防御策略生成和报告合成 4 个智能体，通过黑板架构实现情报共享和协同决策，最后输出结构化结果。后文将逐层细化阐述各核心功能模块的设计原理、实现方案和实验验证。

1.2　ATT&CK 知识图谱构建

为解决传统网络威胁情报中知识碎片化、关系模糊和推理效率低下的问题，本文基于 MITRE ATT&CK 框架[20]，设计有向、类型化、双向的 CTI 本体来构建网络威胁情报知识图谱，为后续的多条路径推理提供支撑。定义知识图谱为四元组 $G=(E,R,T,S)$，其中 E 为实体集合，R 为关系集合，T 为实体类型约束，$S\subseteq E\times R\times T$ 为知识三元组集合。为使推理具备可追溯性和无歧义遍历能力，所有关系均满足类型化与双向可逆双重约束，其类型化约束在关系标签中显式编码目标节点类型，使图执行器在遍历时无须歧义消解即可确定下一跳的目标类型；双向可逆约束要求对任意关系 $r\in R$ 均存在语义一致的有向逆关系 r^{-1}，即 $\forall(h,r,t)\in S\Leftrightarrow(t,r^{-1},h)\in S$，等价于要求关系图在语义上构成对称闭包（symmetric closure）的子集，但保留方向标签以支持从任意节点出发可沿正向或反向边进行多跳推理。

在本体设计中，定义的知识图谱包含以下实体类型：资产、数据源、行动方案（即缓解措施）、入侵集（威胁行为者/APT 组织）、攻击活动等。定义双向关系类型为威胁行为者 $\underset{\text{被使用}}{\overset{\text{使用}}{\rightleftharpoons}}$ 攻击技术等。根据该设计，将已抽取的实体进行关系整理[21]，如将目标实体“APT29”与执行的攻击模式实体“数据混淆”间的关系标记为“利用”，就能构建一

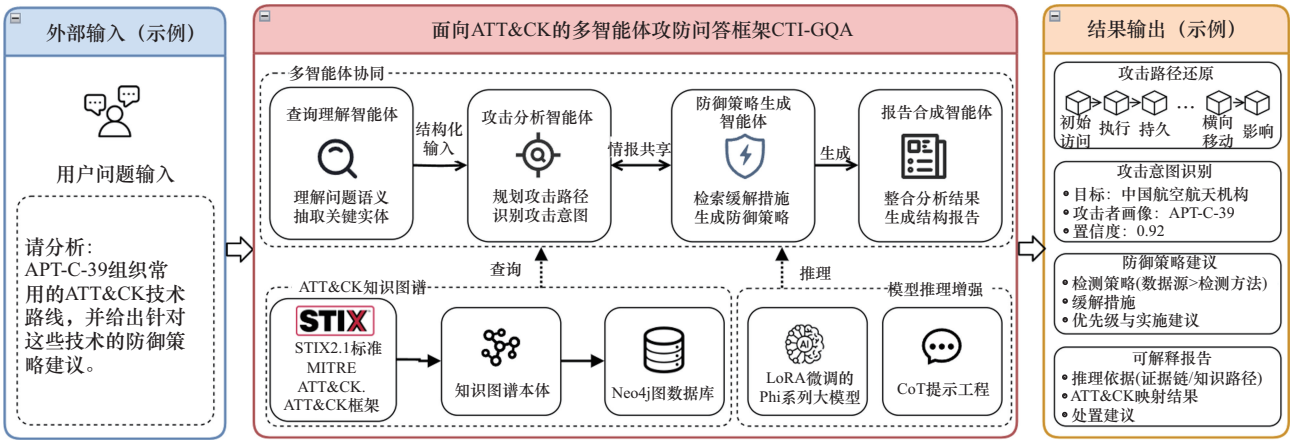

图 1　融合知识图谱与多智能体协同的 ATT&CK 网络威胁情报问答框架

个结构化、易于理解和使用的知识图谱[22]。本体的双向性设计能够支持双向可逆遍历，从而适配攻击链还原、威胁归因等多类CTI任务，但对于超出当前本体覆盖范围的实时情报如NVD漏洞库等，本文在问答框架中设置外部知识库[23]作为过渡，经置信度校验后参与推理，其自动化抽取与本体对齐机制非本文研究重点，暂不展开。基于ATT&CK的威胁情报知识图谱本体如图2所示，其中，每个方块表示一个类别，箭头及文字表示两个类别间的关系。

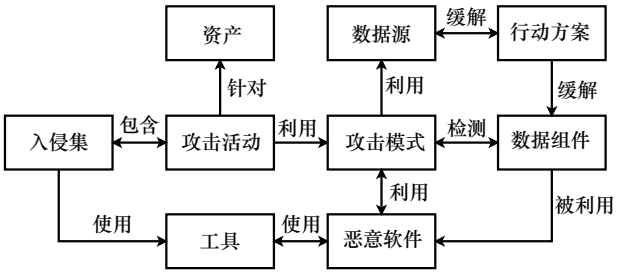

图2　基于ATT&CK的威胁情报知识图谱本体

本文以2026年4月28日发布的MITRE ATT&CK v19.0和STIX 2.1标准为数据源，解析数据包中的领域实体对象（SDO），按类型映射为图节点，并保留STIX标识符与属性作为节点属性来进行实体抽取；解析STIX关系对象（SRO），将其三元组映射为边，并依据目标实体类型对关系标签做类型化重写，从而消除同名关系在不同目标类型下的语义歧义，完成关系抽取，然后对关系进行双向扩展，并进行重复边去重和属性对齐来构建知识图谱，将节点与关系批量导入Neo4j图数据库进行可视化与查询。最终知识图谱共包含883个攻击模式节点、318个行动方案节点、732个恶意软件节点、89个工具节点、31个活动节点、168个入侵集节点、122个数据组件节点、43个数据源节点和14个资产节点，以及它们之间的多种关系，合计2 350个节点、48 795条关系。网络威胁情报知识图谱局部示例如图3所示。

通过将非结构化的网络威胁情报转化为结构化的知识图谱来展示各种威胁实体及其关系，其关系设计为双向类型化，可使模型从任意节点开始进行多跳推理与归因分析。通过此方法，有助于发现潜在的威胁关联，能够为精准应答提供数据支撑，从而提升整体的威胁感知能力和响应效率。

1.3　多智能体协同流程设计

本文所提的网络威胁情报问答框架采用多智能体协同架构，基于LangGraph图状态机实现任务解析、图谱检索、智能体推理和报告合成，完成攻击链还原、意图识别和可解释防御策略生成。多智能体协同流程框架如图4所示。

框架把问答推理过程分为分析、检索、响应3个阶段，多个智能体通过共享状态对象实现数据流转和协同决策。借鉴黑板架构（blackboard architecture）[24]，采用结构化字典存储查询上下文、识别实体、攻击路径、推理置信度、策略覆盖缺口以及反馈信号等关键字段。各智能体以纯函数形式读取所需信息并返回增量更新，由状态机将其合并到全局状态[25]；节点间流转由条件谓词做动态路由，依据置信度阈值或覆盖缺口来决定下一个执行节点[26]。

（1）分析阶段：查询理解与结构化输入生成。接收用户自然语言查询q，由查询理解智能体完成语义解析与实体识别。首先通过实体抽取模块识别威胁实体集合$E=\{e_1,e_2,\cdots,e_n\}$，包含威胁组织、恶意软件、漏洞、攻击技术等核心要素；随后基于预

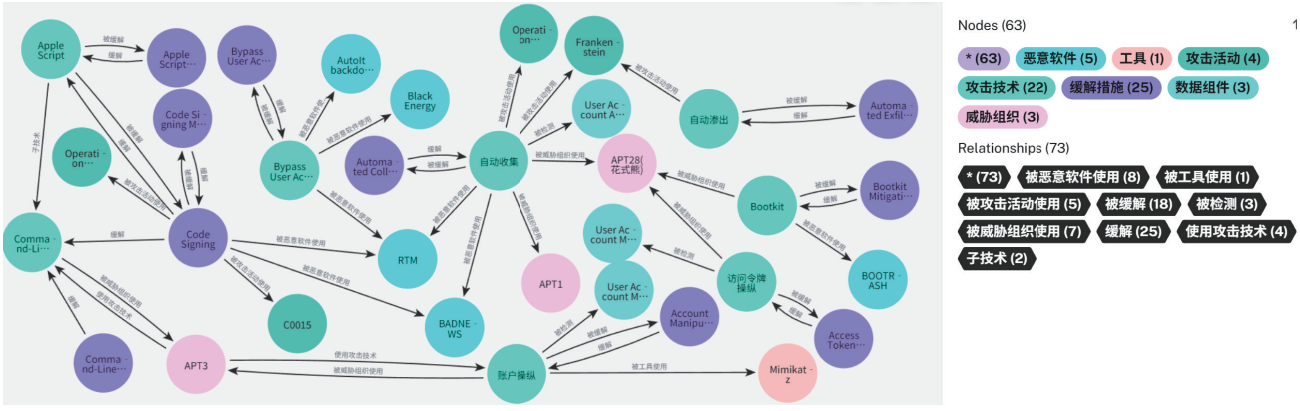

图3　网络威胁情报知识图谱局部示例

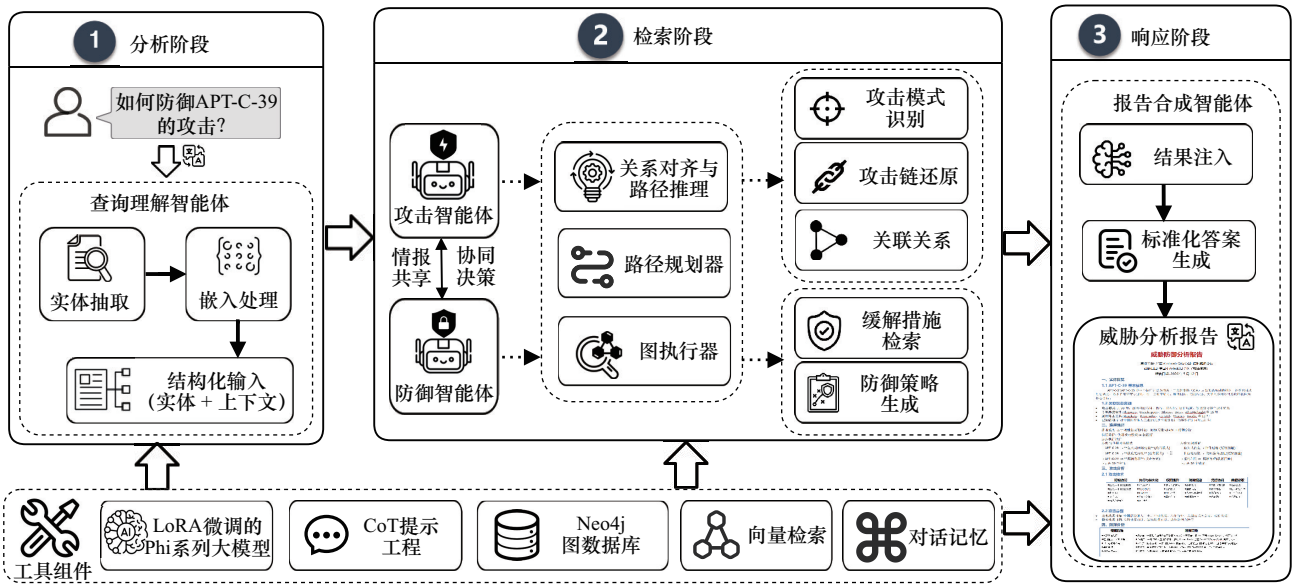

图4　多智能体协同流程框架

训练嵌入模型对实体与上下文进行语义编码，得到嵌入向量 $v_e = \mathrm{Embed}(e)$，并通过多粒度向量对齐将候选实体映射为标准ATT&CK ID，完成自然语言到结构化输入的映射，即：

$$I = \left\{ \left(e_i, v_{e_i}, C \right) \middle| e_i \in E, v_{e_i} = \mathrm{Embed}\left(e_i \right) \right\} \quad (1)$$

其中，I为结构化输入，C为用户查询上下文信息，$\mathrm{Embed}(\cdot)$为BERT类语义嵌入函数。该阶段支持中英双语查询，预处理后的结构化事件序列I作为路径规划器的输入，给后续多跳图谱推理提供基础语义约束[27]。

（2）攻防智能体执行检索阶段：基于LangGraph状态机构建攻击分析智能体和防御策略生成智能体，两者通过情报共享、协同决策实现双向闭环。与单向传递的前馈架构不同，该闭环利用状态对象实现跨智能体反馈，防御智能体找到的策略覆盖缺口可返回到攻击智能体，触发攻击路径的二次补全和重规划[28]。主要通过路径规划和图检索机制完成知识图谱多跳遍历，并基于构建的本体生成带CoT标注的数据集来支撑模型训练与评估。

攻击分析智能体接收结构化输入后，拼接ATT&CK提示模板构建推理指令，让路径规划器生成候选攻击路径集合p，图执行器在知识图谱中对路径p进行合法性校验和多跳遍历，输出攻击推理结果并同步状态：①攻击战术意图，映射至MITRE ATT&CK战术层标签，例如初始访问、持久化等；②时序攻击链，按攻击逻辑进行排序生成TTP序列，如T1003.002→T1027→T1059；③推理溯源证据，记录推理过程依赖的图谱节点和边，实现推理过程可追溯，有效缓解LLM推理时出现的幻觉。

防御策略生成智能体从共享状态对象中读取攻击技术ID集合$T = \{t_1, t_2, \cdots, t_n\}$，采用多级关联匹配策略检索防御实体，检索函数定义为：

$$\mathrm{Retrieve}(t_i) = M_{\mathrm{exact}}(t_i) \cup M_{\mathrm{tech}}(t_i) \cup M_{\mathrm{tactic}}(t_i) \quad (2)$$

其中，M_{exact}、M_{tech}、M_{tactic}分别为精确匹配、关联技术匹配、战术域匹配的缓解措施集合。针对因关系边实测缺失导致检索召回率下降的情况[29-30]，本文通过SP-CoT路径推理挖掘隐式关联来补全缺失边，先以精确匹配为基准，后用关联与战术匹配作为召回增益，降低边缺失影响[31]，并通过反向遍历知识图谱，检查其关联攻击技术是否被攻击分析智能体识别，未识别的写入策略覆盖缺口字段并回流补全。若出现缓解措施冗余、优先级模糊、场景适配性不足等问题，本文微调LLM完成防御策略优化，将检索结果去冗合并，按照危害程度和资产重要性分级，再结合上下文适配具体场景，并基于零信任模型补充相关防御措施。

（3）响应阶段：报告合成与结构化输出。LangGraph状态机完成攻防智能体推理后，报告合成智能体读取整个状态对象，调用LLM对攻击方式、攻击意图和缓解措施进行整合归纳，最终输出结构化的报告内容，包括攻击路径还原、攻击意图识别、防御策略建议。

1.4 思维链提示工程设计与优化

思维链是大语言模型在思考、解决问题时的一系列中间推理步骤，旨在引导大模型循序渐进地思考、解决问题，提高推理能力，使回答更符合预期，保证推理过程的逻辑性和可追溯性。基于这一技术，研究人员不断创新，Wei等[32]首次提出Few-shot CoT，通过推理示例引导模型分步求解；Wang等[33]引入自洽性机制，通过多路径采样与多数投票方法结合来抑制单一路径的随机误差。在结构化知识推理领域，Zhao等[34]提出知识图谱思维链（KG-CoT），将知识图谱中给定实体集合E与关系集合R的多跳路径显式序列化为文本推理链$L=<(e_1,r_1,e_2),\cdots,(e_{n-1},r_{n-1},e_n)>$，其中$e_i\in E,r_i\in R$，从而使模型输出可验证的图谱轨迹。

然而，现有KG-CoT方法多面向开放域知识库，难以满足ATT&CK本体的强类型约束与攻防双向推理需求。为此，本文提出结构化路径思维链SP-CoT，引入类型映射函数$\tau:E\rightarrow T_{\text{type}}$，将实体映射为ATT&CK本体（如$\tau$(APT40)=Intrusion Set），得到类型化路径模板：

$$P_{\text{template}}=\left(\tau(e_1),r_1,\tau(e_2)\right),\cdots,\left(\tau(e_{n-1}),r_{n-1},\tau(e_n)\right) \tag{3}$$

路径规划器的生成目标形式化为在类型约束与关系约束下的最大概率路径搜索：

$$P^*=\underset{P\in P(q)}{\arg\max}\,P\left(P\middle|q,T_{\text{type}},R_{\text{ATT\&CK}};\theta_{\text{LoRA}}\right) \tag{4}$$

其中，T_{type}为类型映射约束，$P(q)$为查询q的候选路径空间，$R_{\text{ATT\&CK}}$为ATT&CK预定义关系集合[35]，θ_{LoRA}为经LoRA[36]微调的模型参数。为进一步抑制单路径生成中可能出现的幻觉现象，借鉴self-consistency[37]的采样聚合思想，对同一查询并行采样N条候选路径$\{P_1,\cdots,P_N\}$，通过多数投票机制确定最终输出：

$$\hat{P}=\arg\max_{P}\sum_{i=1}^{N}\mathbb{I}\left(\text{Sim}\left(P_i,P\right)\geqslant\delta\right) \tag{5}$$

其中，$\mathbb{I}(\cdot)$为指示函数，$\text{Sim}(P_i,P)$为归一化编辑相似度，阈值δ为0.85。若两条路径的归一化编辑相似度≥0.85，则视为等价路径参与投票，避免因LLM同义表述差异导致投票失效。最后通过<PATH>…<SEP>…</PATH>标签显式约束输出格式[38]，使推理链可直接映射为Cypher查询语句，进一步提升路径预测的可靠性与可解释性。面向ATT&CK知识图谱的结构化路径推理提示工程示例如表1所示。

1.5 路径规划与图检索机制

针对网络安全威胁情报领域跨多个实体推理的多跳查询存在的瓶颈问题，本文提出一种融合LoRA微调的路径规划与图检索机制。预训练大语言模型虽依托海量多领域文本展现出通用推理能力，但在CTI等垂直专业场景中性能往往显著衰减，传统的全量微调方法不仅对标注数据规模与质量要求极高，

表1 面向ATT&CK知识图谱的结构化路径推理提示工程示例

设计步骤		设计描述
角色定位		你是一名网络威胁情报分析专家，精通MITRE ATT&CK框架与STIX 2.1标准，能够根据用户问题生成符合知识图谱拓扑结构的可执行关系路径，并基于自我一致性检验确保路径的语义正确性。
任务描述		请从以下网络威胁情报问题中识别实体类型与推理目标，生成一条有效的图谱遍历路径。路径由预定义关系序列组成，输出格式为<PATH>关系1<SEP>关系2<SEP>...<SEP>关系n</PATH>。
补充说明		1. 路径生成需严格遵循知识图谱本体约束，仅使用预定义关系类型； 2. 采用类型占位符机制，路径依赖于节点类型而非特定实体名称； 3. 检查思维的自洽性：①分步推导关系链并进行逻辑检验；②多维度思考列出多种候选路径并交叉验证；③进行一致性校验得出最优路径。
Few-Shot 示例	输入	“APT-C-39针对中国航空航天机构的恶意软件常采用哪些攻击技术?”
	SP-COT	1. 明确实体及类型：“APT-C-39”，实体类型为“入侵组织（Intrusion-Set）”；“中国航空航天机构”，实体类型为“资产（Asset）”。 2. 推导关系及类型：“入侵组织”→“使用”→“恶意软件”（行为关系）；“恶意软件”→“针对”→“中国航空航天机构”（目标过滤）；“恶意软件”→“利用”→“攻击模式”（行为关系）；“攻击模式”→“实现”→“攻击技术（ATT&CK）”（技术映射关系）。 3. 一致性校验：①上下文推理完备，无隐含关系；②经过多维思考，不存在更优的遍历路径；③进行自我校验，符合图谱本体逻辑和路径格式。
	输出	<PATH>使用恶意软件<SEP>针对中国航空航天机构<SEP>采用攻击模式<SEP>实施攻击技术</PATH>

而且算力消耗巨大，还因大幅调整全部参数易引发优化不稳定、梯度异常及过拟合等问题[39-40]。

为此，参数高效微调（parameter-efficient fine-tuning, PEFT）技术被引入，其中LoRA因实现简洁、效果突出而被采用，其核心机制在于对原始权重矩阵 $\boldsymbol{W}_0 \in \mathbb{R}^{d\times k}$ 的更新量约束为两个低秩矩阵的乘积：

$$\boldsymbol{W}_0 + \Delta\boldsymbol{W} = \boldsymbol{W}_0 + \boldsymbol{AB} \tag{6}$$

其中，$\boldsymbol{A} \in \mathbb{R}^{d\times r}$，$\boldsymbol{B} \in \mathbb{R}^{r\times k}$ 且秩 $r \ll \min(d,k)$，$\Delta\boldsymbol{W}$ 表示权重更新。给定输入向量 $\boldsymbol{x}$，前向传播过程为：

$$\boldsymbol{h} = \boldsymbol{W}_0\boldsymbol{x} + \Delta\boldsymbol{W}\boldsymbol{x} = \boldsymbol{W}_0\boldsymbol{x} + \boldsymbol{BAx} \tag{7}$$

在微调过程中，原始权重矩阵 $\boldsymbol{W}_0$ 被冻结，不接受梯度更新，仅对 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 进行更新，使LoRA能够以极低的参数量增量实现高效的领域迁移，显著压缩显存占用与训练耗时。

基于此，路径规划器采用经LoRA微调的Phi系列大模型，以 I 为输入，引入设计的SP-COT提示策略，引导模型显式输出推理中间步骤，继而生成包含实体与关系的结构化路径，如“APT29→使用→战术→目标→资产”。问答框架将该路径转换为Cypher查询，并在知识图谱 G 中执行多跳遍历。在训练过程中，模型需先输出SP-COT思维链，再生成对应的关系路径，从而建立自然语言问题到可执行图查询的映射能力，提升路径规划的准确性与结果可解释性[41]。

图执行器接收路径 p 后进行3类校验：①实体存在性校验，确认源节点和目标节点能够映射到图 G 中的节点；②语义约束校验，确认每条边的起止节点类型符合本体定义；③连通性校验，执行Cypher查询并返回非空路径。若校验失败，触发重规划或退化为单跳检索。

为量化推理路径的可信度，在返回检索结果时，进行置信度计算。对于给定路径 p，其置信度计算采用连乘形式：

$$\zeta = \prod_{(h,r,t)\in P} w(r) \tag{8}$$

其中，关系权重 $w(r)$ 反映该关系在ATT&CK本体中的确定性程度。对于ATT&CK官方预定义的标准关系（如uses、mitigates、detects），$w(r) = 1.0$；对于经LLM抽取并通过人工校验的间接关系（如related to），$w(r) = 0.8$；对于外部知识库补充的临时关系，$w(r) = 0.6$。路径规划与图搜索机制算法如算法1所示。

算法1　路径规划与图执行流程

输入　结构化查询 I，知识图谱 G，LoRA微调语言模型 M_{LoRA}，最大重试次数 $K_{\max}$

输出　检索结果集 R，推理路径 P，溯源证据Trace，置信度 ζ

1)初始化：$R \leftarrow \varnothing$，Trace$\leftarrow\varnothing$，retry$\leftarrow$0
2) while retry < $K_{\max}$ do
3)　$p \leftarrow M_{\text{LoRA}}(I)$// 生成关系路径，形式为 (h_1,r_1,t_1)
4)　for each$(h,r,t) \in P$ do //三元组(源节点，关系类型，目标节点)
5)　　if $(h,r,t) \in G$ then //检查知识图谱 G 中是否存在该三元组边
6)　　Trace $\leftarrow$ Trace $\cup$ { GetProvenance $((h,r,t),G)$ }//记录边信息
7)　　$R \leftarrow$ ExecuteCypher(P,G)//执行合法路径，返回检索结果
8)　　$\zeta \leftarrow$ Confidence(P,G)//计算路径置信度
9)　　return < R,P,Trace,ζ >//
10)　end if
11)　end for
12)　$I \leftarrow$ InjectNegativeFeedback(I,P,Trace)// 负反馈注入，重规划
13) end while
14)$R \leftarrow$ FallbackRetrieval(I,G)//降级为单跳检索
15) return < R,$\varnothing$,Trace,0 >

2　实验验证

2.1　实验设置

为验证CTI-GQA的有效性，本文依托Ubuntu 22.04、CUDA 12.1与PyTorch 2.1的实验环境，设计主观专家评估与客观量化指标两类实验，使用NVIDIA A100 GPU进行模型训练，其中路径规划器以Phi-3为基准模型，采用Unsloth框架进行4-bit量化加载与LoRA监督微调。在微调的参数设置中，训练轮数epoch设为8，初始学习率Learning rate为 3×10^{-4}，序列最大长度max_length为2 048，训练批次Batch_size为8，丢失率Lora_dropout为0.1，放缩因子Lora_alpha为16，秩Lora_rank为16。

主观性能评估方面，邀请10位具有5年以上一

线威胁分析经验的安全专家组成评审组，针对5类典型威胁情报问答任务的测试集，采用双盲方式，隐藏模型身份和输出来源，随机打乱题目和模型顺序，从答案准确率、推理可解释性、信息完整性、策略实用性与术语规范性5个维度根据评分细则进行百分制量化评分。评分汇总前根据Krippendorff's α判定标准[42]和ICC分级标准[43]来对评分进行一致性检验，未达标的予以复核或剔除。

为客观评估模型在威胁情报多跳推理任务上的性能，按照以下方法构造消融实验测试集：依托构建的知识图谱本体，遍历所有满足类型化与双向可逆约束的有效推理关系路径，手动构建700多个带有类型占位符的问题模板，并自动用知识图谱中的实体进行填充；为增加语言多样性并防止规划器记忆模板格式，通过LLM对模板进行释义，最终生成88 209条带CoT标注的样本，按问题所属部分进行分层抽样，按比例划分为训练集74 258条与测试集13 951条。测试集中的节点与问题被重新排列为训练阶段未见的组合，以评估模型的组合泛化能力。另外，为AthenaBench标准化基准测试，以零样本方式向各模型输入AthenaBench中的任务问题，采用该基准6项核心任务，从基础理解、文本推理、空间预测、算法分类、任务达成及答案质量等多个维度量化模型能力。

2.2 问答框架展示

本文构建了面向网络威胁情报领域的CTI-GQA问答框架。该框架选取若干真实APT攻击场景中的典型问题（如防御APT29的攻击）进行测试，分析其给出的答案，发现与通用大模型相比，问答框架在回答相关CTI问题时，不仅能够准确输出涉及的ATT&CK技术编号（如T1133），还能结合CoT进行跳步解释，并在每个推理环节引用对应的检测规则与缓解措施，显著提升了推理过程的透明性与可信度。问答框架应用示例如图5所示。



图5 问答框架应用示例

2.3 实验结果与分析

（1）主观性能评估

本文设计主观性能评估实验，通过横向对比不同模型在威胁情报问答任务中的表现来系统验证CTI-GQA问答框架对基线模型的性能提升效果。选取参数量相近的对照模型进行对比，包括基于LoRA微调的Phi-3-mini-3.8B-Instruct模型、未微调的通用大模型LLaMA-3.2-3B-Instruct、Qwen-2.5-3B-Instruct以及参数量稍大的大模型Mistral-7B-Instruct-v0.3。测试集涵盖五类威胁情报问答任务，测试问题类型如表2所示，共计30道测试题目。专家对模型输出内容进行0～100分量化评分，对评分进行一致性检验，得到各维度Krippendorff's α为0.72和ICC(2,k)为0.82，表明专家评分一致性与可靠性良好，再通过综合加权计算得出最终评估结果，自建测试集模型对比如图6所示。鉴于测试题较小，对CTI-GQA与各基线模型的加权评分差异采用Wilcoxon符号秩检验进行统计显著性检验，CTI-GQA的加权评分为0.003，显著高于各基线模型。

表2 测试问题类型

问题类型	任务说明	题量/道
攻击技术识别	给定恶意软件名称，识别ATT&CK技术编号和名称	6
攻击链还原	根据多阶段证据还原攻击路径和战术阶段	6
攻击意图推断	依据技术序列推断攻击者目标和可能影响	6
防御措施匹配	为技术或攻击链生成缓解、检测和加固措施	6
综合报告生成	生成结构化威胁分析与响应建议	6

由图6可以看出，本文CTI-GQA问答框架在5类问题上的平均得分位居首位，显著高于其他对照模型。在参与评估的通用模型中，Mistral-7B在各任务上表现相对均衡，Qwen-2.5-3B与LLaMA-3.2-3B得分依次降低。对比结果发现CTI-GQA在攻击链路径还原与防御策略匹配两个类别中优势尤其突出。分析原因主要是通用模型在回答路径还原类问题时常出现技术节点遗漏、时序逻辑混乱或幻觉（如编造不存在的ATT&CK技术ID）；而CTI-GQA借助知识图谱的结构化约束与图可执行推理，能够严格沿图谱中的边关系（如“缓解措施关联”）进行跳步回溯，每个输出节点均可追溯到具体STIX

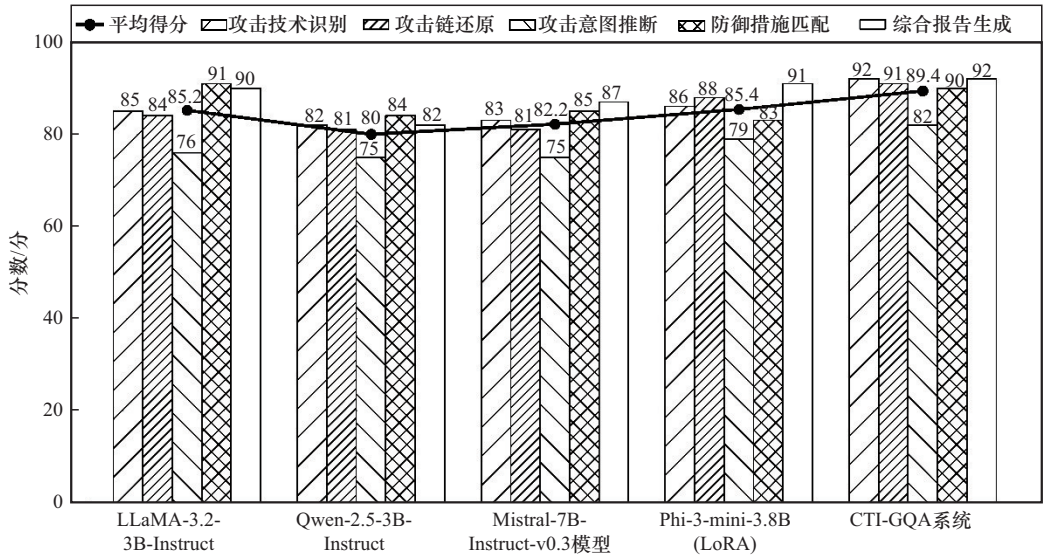

图6　自建测试集模型对比

对象。在攻击意图推断方面，CTI-GQA得分同样高于其他对照模型。专家一致认为，CoT提示与路径规划器提供的分步推理过程能够有效提升意图推断准确性及其可解释性。此外，不含智能体问答框架的LoRA微调Phi-3模型表现与通用模型处于相近水平，说明仅凭微调难以完全弥补威胁情报问答任务中结构化路径推理能力上的不足。

（2）客观性能评估

为了进一步在量化层面验证问答框架性能，本节分别在自建测试集与AthenaBench标准化基准[44]上开展客观评测。

① 消融实验测试。为客观评估模型在威胁情报多跳推理任务上的性能，针对13951条CTI路径推理测试集，采用路径推理准确率以及4项自然语言生成（NLG）通用指标（ROUGE-L、ROUGE-1、BLEU-4、BERTScore-F1），并设置3组模型，包括仅LoRA微调的phi模型、移除SP-CoT的CTI-GQA、完整CTI-GQA来进行消融实验。核心评估指标包括5个指标：路径推理准确率以预测路径与标准答案路径的精确匹配率（EM）[45]衡量多跳推理正确性；ROUGE-L基于最长公共子序列计算生成文本与参考文本的相似度来衡量内容结构完整性；ROUGE-1基于unigram重叠率评估节点覆盖度；BLEU-4基于4-gram匹配度评估词汇准确性；BERTScore利用BERT计算语义相似度来衡量同义改述能力。自建测试集消融实验对比如图7所示，消融实验客观评估指标对比如表3所示。

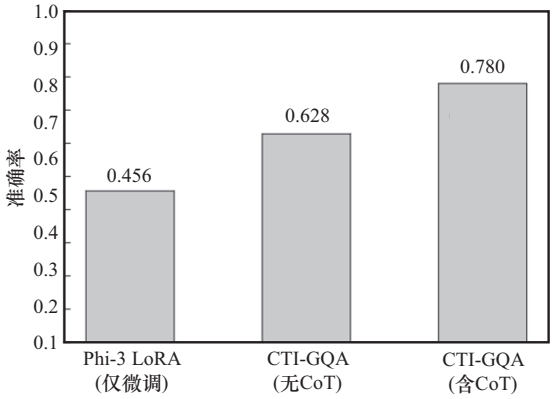

图7　自建测试集消融实验对比

表3　消融实验客观评估指标对比

模型	Path Acc (EM)	R-L	R-1	BLEU-4	BERTScore
Phi-3-mini（LoRA）	0.456	0.530	0.530	0.398	0.808
CTI-GQA（无 SP-CoT）	0.628	0.566	0.567	0.480	0.830
CTI-GQA（含 SP-CoT）	0.780	0.616	0.609	0.574	0.852

由图7可以看出，在攻击路径推理任务中，相较于只进行LoRA微调的基线模型，引入知识图谱与多智能体协同机制后，路径推理精确匹配率由0.456提升至0.628，相对提升37.7%，表明结构化知识约束与多智能体协同可有效增强攻击链推理能力。后加入SP-CoT分步推理，EM指标达到0.780，较无SP-CoT版本相对提升24.2%，验证了结构化分步推理对多跳路径规划的优化作用。

由表3可以看出，在文本生成质量上，CTI-GQA（无SP-CoT）在所有指标上均优于基线模型，其中BLEU-4提升了20.6%，最为显著，表明CTI-GQA框架在威胁情报问答任务中具有良好的语义生成能力。进一步引入SP-CoT模块后，各项指标均有提升，BLEU-4由0.480提升至0.574，相对提升19.6%，这表明SP-CoT通过显式推理过程引导，有效增强了模型对复杂威胁情报问题的语义理解与生成能力。综合来看，完整CTI-GQA在ROUGE-L、ROUGE-1、BLEU-4与BERTScore上均达到最优，证明了本文方法的整体优越性。

上述消融实验结果表明，知识图谱与双智能体能够提供结构化推理约束，缓解模型逻辑混乱问题，并且提出的SP-CoT思维链通过显式分步推理引导模型实现多跳路径规划，抑制知识幻觉，提升推理可解释性。

② AthenaBench标准化基准测试。为确保实验结果的科学性与横向可比性，进一步引入AthenaBench基准测试进行标准化评估。AthenaBench是基于CTIBench[46]构建的CTI分析动态评估基准，涵盖TI知识测试（CKT）、攻击技术提取（ATE）、根本原因映射（RCM）、风险缓解策略（RMS）、漏洞严重性预测（VSP）及威胁行为者归因（TAA）6项任务，其中RMS任务采用F1值（F1-Score）综合评估防御策略的生成质量；其他任务均采用准确率（Accuracy，Acc）作为核心评估指标，各指标取值范围均为0～100%，数值越高表明模型在该维度上的CTI能力越强。为确保评估结果具有全面性和代表性，本文使用公开的AthenaBench-Mini数据集开展对比实验，各模型在该数据集上的CTI任务性能对比如表4所示。

由表4可以看出，在AthenaBench-Mini的CTI任务评测中，本文提出的CTI-GQA在ATE与RMS两类强拓扑关联任务上有显著优势，相较于新提出的SOTA模型Foundation-Sec-8B-Instruct在各方面都有进一步的提升，特别是其ATE准确率达83%，较分数最高模型相对提升12%，RMS的F1值则相对提升30%，验证了知识图谱约束与多智能体协同推理的有效性。但其在CKT、VSP及TAA指标上仍存在性能短板，主要原因分别在于其参数量较小且小模型的跨领域知识有限、缺失漏洞量化的结构化线索及图谱跨活动归因边稀疏的影响，导致其性能未超越其他模型。综上所述，该框架通过融合知识图谱与多智能体协同机制，显著增强了攻击路径规划与防御策略生成的技战术映射准确度，降低了“幻觉”，未来可通过知识蒸馏增强基座、扩充漏洞相关图谱本体等方法，进一步提升其综合性能。

表4　各模型在AthenaBench-Mini上的CTI任务性能对比

模型	CKT (Acc)	ATE (Acc)	RCM (Acc)	RMS (F1)	VSP (Acc)	TAA (Acc)
Qwen2.5-3B	72.60%	4.00%	17.00%	5.58%	60.07%	58.00%
Qwen2.5-7B	19.67%	9.00%	47.50%	6.63%	74.69%	40.00%
DeepSeek-v3.2	87.81%	65.00%	69.50%	23.89%	85.86%	58.75%
DeepSeek-v4-pro	91.00%	79.00%	70.50%	28.15%	88.31%	59.00%
glm-4.7	76.38%	62.00%	68.50%	13.31%	78.58%	51.00%
doubao-seed-2.0-pro	92.00%	74.00%	71.50%	37.30%	81.83%	59.00%
Foundation-Instruct	81.33%	46.00%	54.50%	18.23%	64.59%	56.00%
本文模型	88.33%	83.00%	66.50%	48.50%	80.38%	54.00%

（3）案例分析

为验证本文提出的威胁推理方法在真实网络威胁情报场景中的有效性、规范性与可靠性，基于彭国军等[47]提出的推理攻击方法进行案例分析，选取360安全中心发布的威胁分析报告中的APT-C-39组织的技战术分析以及相关防御策略生成作为验证样本，开展威胁推理实验，本文方法推理结果示例如图8所示。

由图8可知，威胁推理验证实验是以问题“分析APT-C-39组织在攻击过程中利用了哪些ATT&CK技战术，并给出防御策略”为输入，对CTI-GQA框架的推理能力进行验证。本文方法通过深度整合ATT&CK知识图谱与多智能体协同架构，能够精准识别并关联攻击链中的关键技战术且基于检测规则提出相关防御策略。例如，在初始入侵阶段，系统依据语义特征将攻击行为映射至T1195与T1200等标准化战术节点；同时，基于图谱的被缓解关联边，自动推导对应防御措施M1013与M1286，实现从攻击分析到防御策略生成的闭环推理。



图8　本文方法推理结果示例

3　结束语

本文提出了一种基于ATT&CK知识图谱的网络威胁情报问答系统多智能体构建框架。通过构建ATT&CK知识图谱，设计攻击分析和防御策略生成两个专业化智能体，并建立攻防协同闭环机制，实现了从威胁告警到可执行防御的端到端自动化转换。在公开数据集上的实验结果表明，提出的问答框架在攻击技术匹配准确率和缓解措施匹配准确率等关键指标上均优于现有基线方法。消融实验进一步验证了图执行器和RAG检索等核心组件的关键贡献。本文工作为网络威胁情报的自动化分析和智能防御策略生成提供了一种新的技术路径。

本文生成的防御策略主要依赖于预定义的缓解措施和检测规则模板，且威胁知识图谱需随外部情报源持续更新，难以实现实时威胁感知与自主防御。因此，未来工作将从以下两方面展开：一是研究基于强化学习的自适应防御策略学习方法，能够根据实时攻击推理结果动态生成针对性防御方案，提升自主防御能力；二是引入增量知识图谱更新机制，对MITRE官方更新接口、漏洞库等多类异构情报源，进行新增情报清洗、跨源实体对齐、冲突消歧校验、双向关联关系补全，要求内容保持实时同步，并记录图谱更新版本，定期留存全量图谱实体、关系完整图，同步记录节点新增、删除、关系变更的增量日志，支持历史图谱回溯、不同版本推理效果对照，为实现实时威胁感知与自主防御提供支撑。

参考文献：

[1] Psarrou K, Bountakas P, Eleutheriou D, et al. Enhanced-LLM extraction of CTI from unstructured threat reports: a tough nut to crack or a walk in the park? [J]. Future Generation Computer Systems, 2026, 183: 108514.

[2] 涂晨. 威胁情报知识图谱的研究现状与未来趋势[J]. 计算机科学与应用, 2025, 15(2): 153-158.
Tu C. Research status and future trends of threat intelligence knowledge graphs[J]. Computer Science and Application, 2025, 15(2): 153-158.

[3] Noel S, Harley E, Tam K H, et al. CyGraph: graph-based analytics and visualization for cybersecurity[J]. Handbook of Statistics, Elsevier, 2016 (35): 117-167.

[4] Xiang X Y, Ma C C, Zeng L Y, et al. Uncovering multi-step attacks with threat knowledge graph reasoning[J]. Security and Safety, 2025, 4: 2024019.

[5] Li L S, Qiang F, Ma L. Advancing cybersecurity: graph neural networks in threat intelligence knowledge graphs[C]//Proceedings of the International Conference on Algorithms Software Engineering and Network Security. New York: ACM Press, 2024: 737-741.

[6] Bilot T, El Madhoun N, Al Agha K, et al. Few edges are enough: Few-shot network attack detection with graph neural networks[C]//International Workshop on Security (IWSEC). Berlin: Springer, 2024: 257-276.

[7] Cheng Y T, Bajaber O, Tsegai S A, et al. CTINexus: automatic cyber threat intelligence knowledge graph construction using large language models[C]//Proceedings of the 2025 IEEE 10th European Symposium on Security and Privacy (EuroS&P). Piscataway: IEEE Press, 2025: 923-938.

[8] Wang Z H, Fei S Y, Hu Y L, et al. Automated attack knowledge graph construction with large language models[C]//Proceedings of the 2025 2nd International Conference on Computer and Multimedia Technology. New York: ACM Press, 2025: 700-706.

[9] Yang X Z, Zhong R J, Chen Y L, et al. CTI-Thinker: an LLM-driven system for CTI knowledge graph construction and attack reasoning[J]. Cybersecurity, 2026, 9: 106.

[10] Nguyen H C, Tariq S, Chhetri M B, et al. Towards effective identification of attack techniques in cyber threat intelligence reports using large language models[C]//Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25). New York: ACM Press, 2025: 942-946.

[11] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM Press, 2020: 9459-9474.

[12] Houssel P R B, Singh P, Layeghy S, et al. Towards explainable net-

work intrusion detection using large language models[PP]. V1. (2024-08-08)[2026-06-10]. arXiv: arXiv.2408.04342.

[13] Wen H H, Sharma P, Yegneswaran V, et al. MobiLLM: an agentic AI framework for closed-loop threat mitigation in 6G open RANs[C]//Proceedings of the MILCOM 2025 - 2025 IEEE Military Communications Conference (MILCOM). Piscataway: IEEE Press, 2025: 1-6.

[14] Oh H L, Jie J Y J, Siu M L L, et al. Automated post-incident policy gap analysis via threat-informed evidence mapping using large language models[PP]. V1. (2026-01-04)[2026-06-10]. arXiv: arXiv.2601.03287.

[15] Deng G, Liu Y, Mayoral-Vilches V, et al. PentestGPT: evaluating and harnessing large language models for automated penetration testing[C]//33rd USENIX Security Symposium. Berkeley: USENIX Association, 2024: 847-864.

[16] Liu Z. Multi-agent collaboration in incident response with large language models[EB]. (2024-12-17)[2026-07-29].

[17] Wang M J, Dechene R. Multi-agent actor-critics in autonomous cyber defense[EB]. (2026-03-29)[2026-07-29].

[18] Simoni M, Fontana A, Saracino A, et al. TITAN: graph-executable reasoning for cyber threat intelligence[PP]. V1. (2025-10-16) [2026-06-10]. arXiv: arXiv.2510.14670.

[19] Yang L R, Zhu H L, Mu J, et al. KGAgent4CTI: unlocking the power of LLM in threat intelligence[J]. Cybersecurity, 2026, 9(1): 76.

[20] Simoni M, Saracino A. MATRIX: a comprehensive graph-based framework for malware analysis and threat research[C]//Proceedings of the 22nd International Conference on Security and Cryptography (SECRYPT). Bilbao: SciTePress, 2025: 495-502.

[21] 刘子豪, 祝兆鹏, 朱林, 等. 基于知识图谱与大模型的钻井异常智能监测与分析方法[J]. 石油钻探技术, 2026, 54(2): 31-44.
Liu Z H, Zhu Z P, Zhu L, et al. Intelligent monitoring and analysis method for drilling anomalies based on knowledge graph and large language model[J]. Petroleum Drilling Techniques, 2026, 54(2): 31-44.

[22] 赖清楠, 金建栋, 周昌令. 基于大语言模型的网络威胁情报知识图谱构建技术研究[J]. 通信学报, 2024, 45(S2): 33-43.
Lai Q N, Jin J D, Zhou C L. Research on knowledge graph construction technology for cyber threat intelligence based on large language models[J]. Journal on Communications, 2024, 45(S2): 33-43.

[23] Fayyazi R, Trueba S H, Zuzak M, et al. ProveRAG: provenance-driven vulnerability analysis with automated retrieval-augmented LLMs[J]. IEEE Access, 2025, 13: 212815-212826.

[24] Straub J. Modeling attack, defense and threat trees and the cyber kill chain, ATT&CK and STRIDE frameworks as blackboard architecture networks[C]//Proceedings of the 2020 IEEE International Conference on Smart Cloud (SmartCloud). Piscataway: IEEE Press, 2020: 148-153.

[25] Han B C, Zhang S M. Exploring advanced LLM multi-agent systems based on blackboard architecture[PP]. V1. (2025-07-02) [2026-06-10]. arXiv: arXiv.2507.01701.

[26] Salemi A, Parmar M, Goyal P, et al. LLM-based multi-agent blackboard system for information discovery in data science[PP]. V2. (2026-01-31)[2026-06-10]. arXiv: arXiv.2510.01285.

[27] 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. Stroudsburg: ACL, 2019: 4171-4186.

[28] Xu T Y, Zhang D, Mitra K, et al. Verification-aware planning for multi-agent systems[C]//Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg: ACL, 2026: 7528-7546.

[29] Huang Z, et al. Mitigating KG quality issues: a robust multi-hop graphRAG retrieval framework[PP]. V2. (2026-05-19) [2026-06-10]. arXiv: arXiv.2603.14828.

[30] Wei X, Ma J. Design of a graph retrieval-augmented question answering model for the ATT&CK knowledge graph[C]//Proceedings of the 2025 6th International Conference on Computer Engineering and Intelligent Control (ICCEIC). Piscataway: IEEE Press, 2025: 226-230.

[31] Cheng Y T, Bajaber O, Tsegai S A, et al. CTINexus: automatic cyber threat intelligence knowledge graph construction using large language models[PP]. V2. (2025-04-21)[2026-06-10]. arXiv: arXiv.2410.21060.

[32] Wei J, Wang X Z, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]//Advances in Neural Information Processing Systems (NeurIPS). New Orleans: Curran Associates, Inc., 2022: 24824-24837.

[33] Wang X, Wei J, Schuurmans D, et al. Self-consistency improves chain of thought reasoning in language models[C]//International Conference on Learning Representations (ICLR). Vancouver: ICLR, 2023: 1-24.

[34] Zhao R L, Zhao F, Wang L, et al. KG-CoT: chain-of-thought prompting of large language models over knowledge graphs for knowledge-aware question answering[C]//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI). San Francisco: Margan Kaufmann, 2024: 6642-6650.

[35] Strom B E, Applebaum A, Miller D P, et al. MITRE ATT&CK: design and philosophy[R]. MITRE Technical Report, 2018.

[36] Hu E J, Shen Y, Wallis P, et al. LoRA: low-rank adaptation of large language models[C]//International Conference on Learning Representations (ICLR). Vancouver: ICLR, 2022: 1-26.

[37] Francis N, Green A, Guagliardo P, et al. Cypher: an evolving query language for property graphs[C]//Proceedings of the 2018 International Conference on Management of Data. New York: ACM Press, 2018: 1433-1445.

[38] Luo L, Li Y, Haffari G, et al. Reasoning on graphs: faithful and interpretable large language model reasoning[C]//The Twelfth International Conference on Learning Representations (ICLR). Vancouver: ICLR, 2024: 1-24.

[39] Ali I, Subba B. PhishURLDetect: a parameter efficient fine-tuning of LLMs using LoRA for detection of phishing URLs[C]//Proceedings of the 26th International Conference on Distributed Computing and Networking. New York: ACM Press, 2025: 278-279.

[40] Han Z Y, Gao C, Liu J Y, et al. Parameter-efficient fine-tuning for large models: a comprehensive survey[PP]. V7. (2024-09-16) [2026-06-10]. arXiv: arXiv.2403.14608.

[41] Hamzic D, Skopik F, Landauer M, et al. Beyond RAG for cyber threat intelligence: a systematic evaluation of graph-based and agentic re-

trieval[PP]. V1. (2026-04-13) [2026-06-10]. arXiv: arXiv.2604.11419.

[42] Krippendorff K. Content Analysis: an Introduction to Its Methodology[M]. Thousand Oaks: SAGE Publications, 2019.

[43] Koo T K, Li M Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research[J]. Journal of Chiropractic Medicine, 2016, 15(2): 155-163.

[44] Alam M T, Bhusal D, Ahmad S, et al. AthenaBench: a dynamic benchmark for evaluating LLMs in cyber threat intelligence[PP]. V2. (2026-02-14) [2026-06-10]. arXiv: arXiv.2511.01144.

[45] Rajpurkar P, Zhang J, Lopyrev K, et al. SQuAD: 100, 000+ questions for machine comprehension of text[PP]. V3. (2016-10-11) [2026-06-10]. arXiv: arXiv.1606.05250.

[46] Alam M T, Bhusal D, Nguyen L, et al. CTIBench: a benchmark for evaluating LLMs in cyber threat intelligence[C]//Advances in Neural Information Processing Systems 37 (NeurIPS 2024). Vancouver: Curran Associates, Inc., 2024: 50805-50825.

[47] 彭国军, 李家琛, 杨秀璋, 等. 基于大语言模型的威胁情报高效抽取与攻击推理方法研究[J]. 通信学报, 2025, 46(11): 60-72.
Peng G J, Li J C, Yang X Z, et al. Research on efficient threat intelligence extraction and attack inference method based on large language models[J]. Journal on Communications, 2025, 46(11): 60-72.

[作者简介]



郝耀辉（1978– ），女，博士，信息工程大学副教授，主要研究方向为网络信息安全、复杂网络。



李娜（2001– ），女，信息工程大学硕士生，主要研究方向为威胁情报、智能体等。



陈鑫（2003– ），男，信息工程大学硕士生，主要研究方向为智能体、知识图谱等。



秦渝栋（1997– ），男，信息工程大学硕士生，主要研究方向为知识图谱等。



常禾雨（1993– ），女，博士，信息工程大学副研究员，主要研究方向为人工智能安全、网络信息防御。