Computer Science and Application 计算机科学与应用, **2025, 15(2), 153-158**
Published Online February 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.152042

威胁情报知识图谱的研究现状与未来趋势

涂 晨

江西理工大学信息工程学院，江西 赣州

收稿日期：2025年1月15日；录用日期：2025年2月17日；发布日期：2025年2月26日

摘 要

随着物联网的迅速发展，安全问题日益严峻，恶意攻击者通过物联网设备进行攻击，导致数据泄漏。知识图谱通过深度挖掘和关联分析，在安全态势感知和威胁预测中发挥了重要作用。构建物联网威胁情报知识图谱可以整合设备信息、网络流量、攻击事件等多源数据，识别威胁模式，预测潜在攻击路径，实现实时威胁检测与响应。知识图谱的推理能力还能够发现新的威胁关系和异常行为，提升智能防护水平。本文综述从网络安全信息抽取、知识本体建模和图谱存储和推理四个方面展开，探讨了现状和未来趋势，旨在推动知识图谱在物联网安全中的应用。

关键词

威胁情报，知识图谱，实体识别，关系抽取

Research Status and Future Trends of Threat Intelligence Knowledge Graph

Chen Tu

School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou Jiangxi

Received: Jan. 15th, 2025; accepted: Feb. 17th, 2025; published: Feb. 26th, 2025

Abstract

With the rapid development of the Internet of Things, security problems are becoming increasingly severe, and malicious attackers attack through IoT devices, resulting in data leakage. Knowledge graphs play an important role in security situational awareness and threat prediction through deep mining and correlation analysis. Building an IoT threat intelligence knowledge graph can integrate multi-source data such as device information, network traffic, and attack events, identify threat patterns, predict potential attack paths, and achieve real-time threat detection and response. The reasoning capability of the knowledge graph can also discover new threat relationships and abnormal

文章引用：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.
DOI: 10.12677/csa.2025.152042

behaviors, and improve the level of intelligent protection. This paper reviews the current situation and future trends from four aspects: network security information extraction, knowledge ontology modeling, and graph storage and reasoning, aiming to promote the application of knowledge graph in IoT security.

Keywords

Threat Intelligence, Knowledge Graph, Entity Identification, Relationship Extraction

1. 引言

随着信息技术的蓬勃发展，网络承载的信息总量呈爆炸式增长[1]。面对当前互联网上不断增加且呈现出大规模、碎片化的海量信息，传统的搜索引擎技术已无法满足人们准确获取信息的现实需求，这为人们高效、全面地理解数据内涵带来较大的挑战，在人们越来越依赖网络的同时，网络安全问题日益突出。2023 年，广泛存在于许多企业系统中的 Log4j 日志库漏洞被黑客频繁利用，并发动大规模的远程代码执行攻击。2024 年，攻击者发现并利用 Microsoft Exchange Server 的新漏洞，成功入侵了全球数千家企业和政府机构的邮件服务器，窃取了敏感信息并植入后门，导致大规模的数据泄露和系统中断。这些案例显示了随着网络攻击的复杂性不断增强，网络安全的防护变得极为重要。

面对日益复杂多变的攻击手法，传统的安全防护措施难以有效应对新兴威胁，亟需更智能和高效的技术手段来提升网络防护能力[2]。在这种背景下，知识图谱作为一种大规模的安全语义网络，通过从海量分散的网络安全数据中提取实体及其关系，以图结构直观展示概念、属性和语义关联，并整合和分析多源异构的威胁数据，提供了一种安全态势感知和威胁预测手段，通过构建全面的知识网络，不仅能够帮助快速定位威胁的来源与传播路径，还可以通过知识推理挖掘潜在的攻击模式，从而为精准防护和应急响应提供重要支撑，实现实时的威胁检测与响应[3]。

2. 知识图谱及其关键技术

知识图谱的概念源于人工智能和语义网络的发展，其最早可追溯到 20 世纪 90 年代的语义网和本体论的研究，并于 2012 年被 Google 正式推出[4]，作为搜索引擎中一个图结构的知识库。知识图谱作为一种强大的工具，能够通过语义关联和深度挖掘提供新的安全防护视角，将多源异构的威胁数据更好地组织、整合在一起，通过自动化识别和提取关键威胁相关实体，为进一步的安全态势感知和威胁预测奠定基础。但在知识图谱的构建中，离不开精确的信息抽取技术，特别是实体识别和关系抽取[5]，实体识别技术能够从海量的网络安全数据中自动化提取关键的实体信息，而关系抽取则通过识别实体之间的关联构建起图谱的结构化框架，两者相辅相成，共同支持了知识图谱中语义信息的表达和多源数据的融合。因此，信息抽取技术在知识图谱的构建中起到了至关重要的作用，是实现高效威胁预测和响应的第一步。

2.1. 实体识别技术

信息抽取分为命名实体识别和关系抽取。命名实体识别又称实体识别技术，是自然语言处理中的一种关键技术，它通过从文本中自动识别和分类具有特定意义的实体[6]，将识别出的这些实体，进一步与

关系抽取技术相结合，分析实体之间的语义关联和因果关系，从而构建出更具逻辑性和全局视角的威胁情报网络，为攻击路径的溯源、潜在威胁的预测以及安全策略的优化提供强有力的支持。目前常用方法有基于规则匹配、基于机器学习和基于深度学习等。

传统的实体识别方法依赖于规则匹配和字典匹配，它通过定义一组明确的规则或模板，基于模式匹配的方式[7]，从文本中提取符合规则的实体，这种方法简单直接，易于实现，但缺乏灵活性，难以适应未定义的实体或格式不一致的场景。其中，基于规则匹配的方法是命名实体识别研究中最早提出的技术，主要依靠人工制定匹配规则模板和构建专用字典来进行实体的识别和标注。这种方法在识别准确率方面表现优异，但其适用性较差且效率较低，无法处理复杂上下文中的实体。

随着机器学习的发展，基于条件随机场(CRF)和隐马尔可夫模型(HMM)的实体识别方法在特定领域和场景下表现优异，适合对序列数据的精确标注任务。与规则匹配相比，CRF 和 HMM 可以捕捉上下文之间的依赖关系，在序列数据中能够识别复杂的实体边界，提高识别准确率和召回率。然而，由于其依赖特征工程和标注数据的特点，使得性能高度依赖领域知识和数据质量，且需要大量人工标注的语料库作为训练数据，在面对复杂多样化的实体识别任务时，逐渐显现出扩展性和泛化能力的不足。

近年来，基于深度学习的实体识别方法已成为主要的研究趋势，显著推动了自然语言处理和知识图谱构建领域的发展[8]。尤其是循环神经网络(RNN)、长短期记忆网络(LSTM)和 BERT 等预训练语言模型，适合处理含有复杂上下文依赖关系的任务，如跨句子级别的实体识别、文档级别的关系提取。深度学习方法能够通过训练自动学习语料中的语义特征和模式，极大减少了对人工干预的依赖，且可以在较少标注数据的情况下适配新的领域任务，显著提升了实体识别的准确性。此外，多模态实体识别通过结合文本、图像和语音等多种数据源，进一步拓宽了应用场景，能够更全面地识别实体并提高识别效果。然而，深度学习模型的训练和推理需要大量计算资源，在数据不足的场景中，模型性能可能无法充分发挥。与基于规则或机器学习的方法相比，深度学习模型的决策过程较难解释，尤其在安全和关键领域中可能面临信任问题。

2.2. 关系抽取技术

关系抽取技术旨在从文本中自动识别和提取实体之间的语义关系，从而生成<E, R, E>形式的实体关系三元组[9]。这些关系三元组是构建知识图谱的基本单元，用于表示实体及其之间的逻辑和语义关联。通过三元组的形式，能够将分散的信息结构化地存储在知识图谱中，从而为数据的组织、可视化和推理提供基础。其中，E 表示实体，如“样本”、“病毒”、“漏洞”、“攻击类型”等，R 则为实体之间的关系，如“属于”、“关联”、“攻击”、“利用”等。在构建威胁情报知识图谱的过程中，关系抽取技术至关重要，它不仅连接了通过实体识别技术识别出的关键实体，还揭示了这些实体之间的动态交互和关联模式，这对于描述复杂的网络攻击、恶意行为、漏洞利用等动态关系至关重要。关系抽取技术在威胁情报知识图谱构建中主要包括基于规则的方法、基于机器学习的方法和基于深度学习的方法、基于远程监督的方法和基于开放信息提取的方法。

基于规则的方法通过预定义规则或模板，从文本中提取实体之间的关系，适合特定领域的小规模任务，如检测特定格式的攻击路径或漏洞利用描述。但这种方法依赖领域专家的知识和固定的模式，且难以适应复杂的文本和多样化的关系表达，扩展性差。基于机器学习的方法利用有标注的训练数据，通过机器学习模型(如支持向量机、逻辑回归)预测实体之间的关系。它适合数据量较大的威胁情报任务，如挖掘漏洞描述报告中的关联关系。缺点是需要人工标注数据作为训练集、依赖特征工程，且效果受特征质量影响较大，难以处理复杂上下文等。基于深度学习的方法是通过神经网络模型如卷积神经网络、循环神经网络、BERT 等自动学习语料中的特征，提取实体之间的关系，此方法无需手动设计特征就能捕捉复

杂的语义和上下文信息，适合处理非结构化大规模数据、长文本和上下文依赖关系较强的场景。但深度学习需要大规模标注数据和高计算资源，占用内存过高。基于远程监督的方法是通过对齐已有的知识库和未标注文本，自动生成标注数据用于训练关系抽取模型，该方法无需人工标注大规模数据，能快速生成大量训练样本，适用于领域知识库丰富但人工标注成本高的场景。但生成的标注数据可能存在噪声，影响模型性能。基于开放信息提取的方法通过从文本中直接抽取所有可能的实体关系，生成关系三元组，信息提取过程中通常不依赖标注数据或预定义规则，能处理开放领域的文本，适用于动态威胁情报中挖掘潜在未知关系。但其提取的关系质量较低，可能需要后续人工筛选。

基于深度学习和远程监督的方法因其自动化特性和泛化能力，正逐渐成为主流。综合运用多种方法，如混合技术或集成方法，可以在实际应用中实现更高效的关系抽取，提升威胁情报知识图谱的构建效果，推动其在网络安全中的广泛应用与发展。

关系抽取是构建威胁情报知识图谱的重要技术，它能够从大量非结构化文本中提取出实体之间的关系信息，形成有价值的三元组。关系抽取不仅帮助知识图谱从零散的信息中提取出有意义的关联，还能够为后续的攻击路径分析、风险评估和态势感知提供支持。通过关系抽取，威胁情报知识图谱能够映射出攻击者的行为、攻击目标、漏洞利用等之间的关联，为网络安全防御提供更加清晰的视图。随着自监督学习、多模态学习和跨领域迁移的进展，关系抽取将在网络安全等领域发挥越来越重要的作用[10]。

2.3. 知识图谱推理

知识图谱推理是通过分析知识图谱中的实体、关系及其结构，推导出新的知识或信息的过程[11]。它通过不同的推理方式，包括路径推理、规则推理和模式推理，利用图卷积网络(GCN)、知识图谱嵌入、逻辑推理等技术，对图中的信息进行深度挖掘和推导。路径推理可以帮助发现实体间潜在的间接关系，规则推理则通过已定义的逻辑规则推导出新的事实，而模式推理通过已知的关系模式进行预测和推导。知识图谱推理广泛应用于智能问答、推荐系统、疾病诊断、网络安全等多个领域，能够有效地发现隐藏的知识、预测未来的趋势、识别复杂的关联模式，并辅助决策。

尽管知识图谱推理能够在许多应用场景中提供有价值的信息，但它也面临一些挑战。首先，随着知识图谱规模的不断扩大，推理的计算复杂度也会急剧增加，这对实时推理提出了高要求[12]。其次，推理结果的透明性和可解释性是深度学习和图推理中的一个难题，尤其是在关键领域如医疗、金融中，推理过程和结果的解释显得尤为重要。未来，知识图谱推理将朝着更加高效、智能和自动化的方向发展。随着计算能力的提升和新型算法的应用，推理效率将大大提高，特别是在大规模知识图谱的实时推理中。此外，推理系统将更加注重可解释性和透明度，尤其在医疗、金融等高风险领域，以增强信任度和可验证性。跨领域的推理技术也将成为重要发展方向，通过融合多个领域的知识图谱，提供更加全面的决策支持，促进各行业的数字化转型和创新。

2.4. 网络安全知识本体构建

本体构建是知识图谱的基础，旨在为知识图谱提供统一的语义框架[13]。本体通常定义了领域内的基本概念、实体类别及其之间的关系，确保知识图谱中的所有实体和关系都遵循一致的语义规则，旨在系统化和结构化地定义和组织网络安全领域的核心概念及其相互关系，为威胁情报知识图谱提供了一个统一和标准化的知识框架，促进了多源异构数据的有效整合和语义关联。

通过本体建模，组织能够系统地组织和管理大量分散的安全数据，提升信息检索和智能分析的效率，准确的本体模型有助于实现知识推理和自动化决策支持，增强威胁检测、事件响应和风险评估的智能化水平。进行本体构建时，首先，需明确主要实体类别，如“威胁行为者”、“恶意软件”、“漏洞”、“攻

击手法”和“受害资产”等，并为每类实体设定属性以描述其特征。其次，定义实体之间的语义关系，例如“使用”、“利用”、“攻击”、“影响”等，建立实体关系三元组<E, R, E>，以准确反映实际威胁场景中的关联。为了确保本体的语义一致性和互操作性，通常参考和整合国际标准如 STIX 和 MITRE ATT&CK 框架[14]。此外，本体建模过程中需考虑知识的可扩展性和动态更新能力，以适应不断变化的威胁环境。最终，通过精确的本体建模，威胁情报知识图谱能够实现多源数据的深度融合和语义关联，提升威胁检测、事件响应和风险评估的智能化水平，助力网络安全防护的全面提升。

随着知识图谱应用场景的多样化，如何构建灵活、动态、可扩展的本体变得越来越重要。例如，针对不同领域(如医疗、金融、法律等)的本体，如何将领域知识与通用知识结合，实现跨领域的知识共享和语义互操作性，是当前的一个研究方向。在数据隐私和安全问题日益重要的今天，如何通过隐私保护技术(如差分隐私、同态加密)保护在知识图谱构建过程中涉及的敏感数据，也是未来技术发展的关键之一。

2.5. 知识图谱存储

知识图谱存储是将知识图谱中的实体、关系和属性等结构化数据进行持久化管理的过程，旨在高效处理和查询大规模知识图谱数据，实现快速检索、关联和推理[15]。因此，通常采用专门设计的图数据库和存储系统，如 Neo4j、Apache Jena TDB 和 RDF4J，这些工具支持图结构数据的高效处理、语义查询和复杂的图查询操作，并兼容 RDF 和 SPARQL 等标准化语义数据存储与查询语言。知识图谱存储需要在性能、可扩展性和查询能力之间取得平衡，以满足大规模和复杂应用的需求。不同的存储工具在性能、扩展性和推理能力方面各具特色，适用于不同规模和复杂度的知识图谱应用，确保数据的高效管理和智能分析。

随着知识图谱在金融、医疗、政府等行业的广泛应用，数据一致性和安全性成为存储系统设计中的核心问题。在分布式知识图谱存储中，数据一致性尤为关键。为了确保跨节点操作的一致性，通常需要在强一致性和最终一致性之间进行权衡。对于金融、医疗等高要求场景，则需采用强一致性模型，确保系统在任何时刻的数据状态保持同步。分布式事务管理，如两阶段提交协议(2PC)和三阶段提交协议(3PC)，能有效保证跨多个节点的数据一致性，同时通过锁机制和版本控制避免并发操作中的冲突。数据安全性和隐私保护也是知识图谱存储系统中亟待解决的问题，尤其是在涉及敏感数据时。

知识图谱存储系统的改进将集中在提升数据一致性、扩展性和安全性的综合能力上。随着数据量和应用场景的不断增长，存储系统需要在处理大规模图数据时，优化分布式存储架构，提高查询效率和系统响应能力。对于一致性问题，采用更灵活的分布式事务协议和一致性模型，结合新兴的技术如区块链和多版本控制，将是未来发展的趋势。同时，随着数据隐私保护法规的不断完善，知识图谱存储系统需要更智能的隐私保护技术，如同态加密和可验证计算，来确保敏感数据的安全性。在智能化和自动化方面，结合 AI 和机器学习技术，未来的知识图谱存储系统将能够自适应不同应用需求，自动优化存储和安全策略，提升数据管理的效率和精准度[16]。

3. 总结

威胁情报知识图谱的未来发展将聚焦于提升其智能化、动态化和安全性，以更好地应对日益复杂的网络威胁环境。首先，多源异构数据的高效融合与处理仍是关键难点，需要研发先进的自动化数据清洗、转换与整合技术，以及满足大规模数据处理需求的存储与计算架构。研究将致力于开发更加先进的数据融合技术，使其能够高效整合来自物联网、云计算和大数据等多源异构的数据，确保知识图谱的全面性和准确性。其次，实体识别与关系抽取的准确度和效率亟待提升，需要充分利用自然语言处理和机器学习等前沿工具。再次，确保知识图谱本身的安全性也是不可忽视的研究方向，需要建立防篡改、强化访

问控制与实施安全审计等机制，防止图谱被恶意攻击和篡改。标准化和互操作性的研究将促进不同系统和组织之间的威胁情报共享，推动国际合作与信息互通。最后，需要增强知识图谱的动态更新能力，提高知识图谱的自动化水平，减少对人工干预的依赖，并开发有效的评估和验证方法，以确保其有效性和可靠性。通过结合图神经网络和深度学习等先进技术，进一步提升图谱在威胁检测、风险评估和事件响应中的智能化水平，以适应快速变化的威胁环境，研究将探索实时数据流处理、增量学习和在线学习策略。通过这些研究方向的持续推进，物联网威胁情报知识图谱将在网络安全领域发挥更为显著的作用，提供更高水平的智能防护，全面提升组织的风险防范能力和安全响应效率。

参考文献

[1] 孙威. 面向网络威胁情报的知识图谱构建技术研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2023.

[2] 付瑞. 基于实体关系联合抽取的领域知识图谱构建与应用[D]: [硕士学位论文]. 昆明: 云南大学, 2022.

[3] 李涛. 威胁情报知识图谱构建与应用关键技术研究[D]: [博士学位论文]. 郑州: 战略支援部队信息工程大学, 2020.

[4] 董聪, 姜波, 卢志刚, 等. 面向网络空间安全情报的知识图谱综述[J]. 信息安全学报, 2020, 5(5): 56-76.

[5] 王通, 艾中良, 张先国. 基于深度学习的威胁情报知识图谱构建技术[J]. 计算机与现代化, 2018(12): 21-26.

[6] 张舒越, 詹昊谋, 李昕泽, 等. 网络安全知识图谱构建与应用研究综述[J]. 网络空间安全科学学报, 2024, 2(3): 79-106.

[7] 赵弘扬. 基于知识图谱的网络威胁情报信息抽取技术研究[D]: [硕士学位论文]. 广州: 广东技术师范大学, 2024.

[8] 孙艺凡. 基于知识图谱的网络安全威胁情报推理技术研究[D]: [硕士学位论文]. 北京: 华北电力大学, 2024.

[9] 朱鹏程. 面向威胁情报的命名实体识别方法研究[D]: [硕士学位论文]. 广州: 广州大学, 2024.

[10] 任乐, 张仰森, 刘帅康. 基于深度学习的实体关系抽取研究综述[J]. 北京信息科技大学学报(自然科学版), 2023, 38(6): 70-79, 87.

[11] 李序, 连一峰, 张海霞, 等. 网络安全知识图谱关键技术[J]. 数据与计算发展前沿, 2021, 3(3): 9-18.

[12] 尚文利, 朱鹏程, 王博文, 等. 面向威胁情报的知识图谱构建关键技术[J]. 自动化博览, 2023, 40(1): 15-19.

[13] 史慧洋, 魏靖烜, 蔡兴业, 等. 威胁情报提取与知识图谱构建技术研究[J]. 西安电子科技大学学报, 2023, 50(4): 65-75.

[14] 崔孟娇, 姜政伟, 陈奕任, 等. 面向威胁情报的大语言模型技术应用综述[J]. 信息安全学报, 2024, 9(5): 1-25.

[15] 张玉臣, 孙澄, 姜迎畅, 等. 融合威胁情报与知识图谱的网络攻击溯源方法[J]. 情报杂志, 2024, 43(8): 72-83, 91.

[16] 李昌建, 于晗, 陈恺, 等. 物联网威胁情报知识图谱综述[J]. 网络空间安全科学学报, 2024, 2(2): 18-35.