王瀚昀, 刘座臻, 张子赫,等. 基于知识图谱的铁路网络安全领域资产测绘与应用研究[J]. 智能计算机与应用,2026,16(7):154-160. DOI:10.20169/j.issn.2095-2163.24110101

基于知识图谱的铁路网络安全领域资产测绘与应用研究

王瀚昀, 刘座臻, 张子赫, 马思宇, 李钊洋
(中国铁路哈尔滨局集团公司 信息技术所, 哈尔滨 150006)

摘 要: 在铁路信息化建设向数字化转型的大背景下,网络安全领域资产的智慧化管理和应用成为目前发掘资产潜在价值、提升资产管理质量的必由之路。在阐述铁路网络安全领域信息资产管理现状的基础上,将资产信息与安全数据紧密结合,构建出了铁路网络安全领域信息资产知识图谱,并基于该图谱设计出了完整的应用架构与应用场景。为期半年的实践结果表明,构建的知识图谱能够显著提高运维人员获取目标网络设备及其相关信息的速度,对于提高资产管理水平和漏洞风险防控能力也提供了有力支持,具有一定的推广和应用价值。

关键词: 知识图谱; 图数据库; 知识抽取; 资产管理; 数字化转型
中图分类号: TP393 **文献标志码**: A **文章编号**: 2095-2163(2026)07-0154-07

Research on asset mapping and application in railway network security field based on knowledge graph

WANG Hanyun, LIU Zuozhen, ZHANG Zihe, MA Siyu, LI Zhaoyang
(Information Technology Institute, China Railway Harbin Bureau Group Co., Ltd., Harbin 150006, China)

Abstract: In the context of the digital transformation of railway information construction, the intelligent management and application of assets in the field of network security have become an important way to explore the potential value of assets and improve the quality of asset management. On the basis of elaborating on the current situation of railway information asset management, this paper closely integrates asset information with security data, constructs a knowledge graph of information assets in the field of railway network security, and designs a complete application architecture and application scenarios. After half a year of exploration and practice, it has been shown that the knowledge graph constructed in this paper can significantly improve the speed of obtaining network devices and their related information, and provide strong support for improving asset management and vulnerability risk prevention capabilities, which has certain promotion and application value.

Key words: knowledge graph; graph database; knowledge extraction; asset management; digital transformation

0 引 言

近年来,网络环境和攻击方式的复杂多变给铁路安全工作带来了新的挑战。作为铁路网络安全工作的核心部分,信息资产安全一直是相关部门关注的重点。然而随着资产数量的激增以及资产类型的多样,铁路网络安全领域资产管理仍然面临着资产台账不清晰、安全标准不统一、责任链条不完整、态势感知不敏锐等问题,难以进行有效的整合和分析,资产管理的效率和准确性都受到了严重影响,给网络安全带来了潜在的风险[1-7]。

作为目前较为主流的一种知识表示方式,知识图谱以数据为驱动,通过<实体,属性,关系>三元组的形式分析、挖掘各个实体之间的关联关系,在数据分析、智能问答、风险评估、知识管理等领域均被广泛引用[8-15]。通过构建铁路网络安全领域信息资产知识图谱,实现了铁路信息资产自动化、精细化、智能化管理,详细展示了信息资产之间、资产与安全漏洞之间的关联关系,大大提高了资产管理的效率和质量,为路局集团公司资产分析、安全防护以及管理决策提供了重要的支撑和保障作用。

1 网络安全领域信息资产管理现状

在网络安全领域,信息资产管理是一项复杂而至关重要的任务,涉及对硬件、软件、数据和服务的识别、分类、监控和保护,尤其在铁路行业,信息化资

作者简介: 王瀚昀(1993—),男,硕士,工程师。主要研究方向:应用软件研发,计算机应用技术。Email:championericwhy@qq.com。

收稿日期: 2024-11-01

产种类繁多、数量巨大，使得信息资产管理面临着以下 3 个方面的主要问题[16-22]：

（1）资产识别与状态管理不精确。由于技术或管理的原因，部分信息资产长期处于未被发现或登记的状态，信息资产的状态和配置信息也尚未具备实时更新的能力，导致资产管理单位的安全资源无法根据资产的重要性和敏感度进行合理分配。

（2）资产管理与安全防护不匹配。通常情况下，资产管理涉及业务主管单位、运维单位和研发单位等诸多部门，铁路企业组织架构的特殊性使得各部门之间的协同手段匮乏，企业资源和信息难以实现有效共享，资产管理与安全策略之间往往缺乏紧密的联动，从而降低了安全防护的整体效率和效果。

（3）资产监护与风险预警有缺失。在资产管理的过程中缺乏有效的预警机制，致使资产管理单位未能提前识别和防范网络安全事件导致的关键业务系统暂时或永久性中断，尤其面对复杂的网络攻击，无法有效评估和管理相关风险，缺乏必要的准备和应对措施，一定程度上限制了资产信息数据的管理与利用。

2　网络安全领域信息资产知识图谱关键技术

2.1　网络自动拓扑发现与构建技术

网络自动拓扑发现与构建技术是网络安全领域信息资产知识图谱绘制的关键，典型的网络自动拓扑构建方法有 Traceroute 拓扑测量法、SNMP 协议拓扑测量法以及 LLDP 协议拓扑测量法等。其中，Traceroute 拓扑测量法能够快速定位网络数据包的传输路径；SNMP 协议拓扑测量法在网络设备和链路状态监控方面占据优势；LLDP 协议拓扑测量法在设备探测的基础上可以展示设备间的连接关系映射。利用网络自动拓扑发现与构建技术，通过对网络设备的配置信息、性能数据等信息进行自动检测，能够形成当前网络环境中所有信息资产的要素实例及其关联关系，进而辅助知识图谱数据层构建。

2.2　漏洞扫描和安全事件预警技术

漏洞扫描和安全事件预警技术是网络安全领域信息资产知识图谱实现漏洞精确定位和安全监控预警功能的核心，依托诸如 OpenVAS、Snort、TheHive 等成熟的网络安全工具，利用对现有网络环境中的所有信息资产开展动态、静态应用程序安全测试等漏洞扫描检测手段，强化知识图谱的漏洞扫描和网络安全预警能力，使知识图谱智能地关联漏洞信息、资产数据和威胁情报，并为漏洞修复和安全加固提供优先级建议。

2.3　自然语言处理技术

自然语言处理技术是网络安全领域信息资产知识图谱构建的基石，按照图谱构建流程可将自然语言处理技术划分为实体识别技术、关系抽取技术以及文本分类技术。其中，实体识别技术能够自动地从大量非结构化数据（如安全报告、论坛帖子、新闻文章等）中准确地识别出人名、地名、机构名等实体；关系抽取技术能够抽取不同实体间的各种关系，并将这些信息整合到知识图谱中，显著增强知识图谱的语义丰富度；文本分类技术能够将抽取到的有用信息自动归类到不同的主题或类别中，如漏洞类型、攻击手段或安全事件等级等，便于知识图谱更好地组织和检索相关信息。

3　网络安全领域信息资产知识图谱构建

3.1　构建流程

传统的知识图谱构建方式主要有自顶向下和自底向上两种。其中，自顶向下的构建方式利用本体描述语言确定图谱的构建范围以及数据的组织形式，再通过实例填充操作得到目标知识图谱，适用于已有固定体系的知识领域；而自底向上的构建方式则是以底层数据为基础，通过知识抽取方法提炼出基础三元组要素再进行概念层次的划分，最终完成知识图谱构建，适用于体系概念模糊的知识领域。考虑到网络安全领域知识图谱的深度和专业性，在图谱构建的过程中将上述 2 种方法相结合，利用自顶向下的方法统一对网络安全领域资产实体的类型、属性以及关联关系进行标准化和规范化的定义，然后依靠丰富的信息化资产数据，自底向上提取实例要素映射到概念层次领域，实现图谱模式层与数据层之间的优化，具体流程如图 1 所示。

3.2　本体定义

作为知识图谱构建的首要环节，本体定义是定义知识单元、单元属性和单元关系，形成知识图谱本体模型的过程。网络安全领域信息资产知识图谱的本体，记录了信息资产中涉及的实体概念及其相互关系，是对信息资产特性的形式化描述。综合分析领域内相关概念体系和业务知识，本体要素包括：

（1）本体单元，即网络安全领域信息资产所有实体的集合，包含人员、应用系统、漏洞、虚拟机、网闸、交换机、防火墙、操作系统（Windows、Linux 等）、中间件（Nginx、Redis 等）等。

（2）本体属性，即各类实体的属性集合，包含姓名名称、品牌名、IP 地址、版本号、所属单位、端口、

主机名、机房、机柜、网络域等。

(3)本体关系,即各类实体的关系集合,包含运维、安装、所属、运行、接入等,目前图谱总计 22 类本体单元、60 类本体属性、22 类本体关系,其主要内容如图 2 所示。

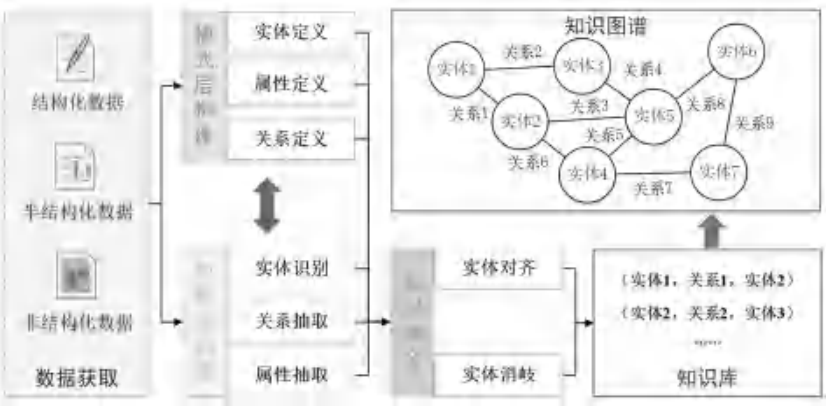


图 1 知识图谱构建流程图

Fig. 1 Process diagram for constructing knowledge graph

本体要素	名称	描述
实体	小型机	为领域内应用系统,虚拟机等提供基础支持的小型机,通过品牌、实例描点等进行区分
	应用系统	领域内运行的应用系统,包括应用的业务范围,部署的机器(虚拟机)等
	人员	应用系统和设备的运维人员和责任人员等
	漏洞	数据库、操作系统及中间件等已知的安全漏洞
	操作系统	硬件服务器和虚拟机服务器所安装的操作系统,通过系统类型、版本号等进行区分
	中间件	为各应用系统提供服务的中间件,包括Redis、Nginx、Elasticsearch等
	存储	领域内采用的实体存储,通过品牌、型号、容量等进行区分
属性	中间件属性	名称、端口、安装路径、版本号、IP等
	人员属性	姓名、单位、部门、联系方式等
	漏洞属性	漏洞编号、漏洞名称、漏洞描述、危险等级等
	操作系统属性	名称、IP、网络域、版本号等
	应用系统属性	名称、业务描述、建设性质、等保等级、备案号、部署方式等
关系	运维	描述了人员与硬件设备、应用系统、中间件之间的联系,如:张三运维A系统
	安装	描述了中间件与操作系统或者操作系统与虚拟机之间的联系,如:虚拟机A安装了B操作系统
	所属	描述了人员与部门之间的联系,如:张三所属于A部门
	运行	描述了应用系统与虚拟机或小型机之间的联系,如:A系统运行在B虚拟机
	接入	描述了交换机、网络等网络设备之间的联系,如:A接入交换机接入B三层交换机

图 2 网络安全领域信息资产主要本体要素示意图

Fig. 2 Schematic diagram of the main ontology elements of information assets in the field of network security

3.3 知识抽取

知识抽取是图谱质量的关键保证,其目标旨在精准抽取出本体定义环节中定义的实体、属性以及关联关系。知识图谱的数据来源大致分为 3 类。第 1 类是铁路信息系统配置库、铁路信息装备资源管理台账、铁路网络安全标准及制度等结构化数据;第 2 类是漏洞扫描报告、资产探测报告等具有一定结构特征的半结构化数据;第 3 类是铁路信息化资产地理数据等没有固定格式且数据量大的非结构化数据。

(1)对于组织良好、易于访问的结构化数据,由于其具有明确的字段、属性和关系定义,可直接从表格或关系型数据库向图数据库中映射,以获得构建项目图谱所需的实体名称、属性信息、关联关系。

(2)对于具有一定结构特征的半结构化数据,可利用数据中的结构化的标签,将标签中的内容按非结构化文本进行处理,或者利用爬虫、正则匹配、文本挖掘、机器学习等实体识别技术对其实体以及属性进行识别。

(3)对于没有固定格式且数据量大的非结构化数据,可采取如下的抽取流程:首先,使用 Scrapy、BeautifulSoup 等 Python 库对原始语料进行自动抓取;其次,利用正则表达式对抓取的文本数据进行规范化处理和初步清洗,去除文本中非语义表达的中英文字符、HTML 标签、URL 链接、文件路径及代码等非文本内容等;接着,通过 jieba 分词的自定义词

典数据进行基于词典的分词；最后，对处理后的数据进行词性标注、实体识别、关系识别等处理。

3.4　知识融合

考虑到网络安全领域信息资产数据集来源广泛，同一实体或属性可能存在多种表述方式，存在一定的歧义与冗余，需要利用知识融合技术对抽取的数据做进一步清洗。知识融合是把多源异构数据整合到一个统一的知识框架中，解决一词多义问题的技术。基于领域内相关概念体系建立了信息资产原始语料库，将知识抽取获得的实体数据与关系数据作为待消歧实体，利用 Dice 系数来计算实体间的文本相似度，Dice 系数越大，2 个实体间的文本相似度越高。对于 Dice 系数较低的待消歧实体，可对其进行规范化处理后作为新增语料加入语料库，从而提高知识融合过程中抽取数据与语料库之间映射的准确性。

3.5　知识存储

知识图谱的存储技术主要包括 RDF 存储、关系型数据库、图数据库存储。这 3 种方式针对不同的使用场景各有优劣。由于网络安全领域资产知识图谱所包含的实体数量多，实体间的关系复杂，因此选用时下流行的 Neo4j 图数据库来完成“实体-关系-实体”三元组的存储工作。

Neo4j 图数据库支持多种形式的数据导入，包含：LOAD CSV 命令、APOC 拓展库、Java(Python)等编程语言编写脚本、neo4j-admin 工具及第三方应用导入。采用 neo4j-admin 工具的方式通过命令“neo4j-admin load --database=数据库名称 --from=CSV 文件路径 --force”将前文产生的大量三元组数据导入 Neo4j 数据库中。数据导入完成之后即可通过编写 Cypher 语句进行图谱的查询，例如：“MATCH (n:system)-[p]->(r) RETURN n, p, r LIMIT 25”语句可以查询出由 system 实体延伸出的前 25 条三元组。

4　网络安全领域信息资产知识图谱应用

4.1　应用架构

为了能够更好地发挥图谱的价值，设计了完整的应用架构，如图 3 所示。

(1)统一门户：基于 Model-View-ViewModel 框架为用户提供统一登录门户，并采用 Web Service 架构与其他系统进行信息交换。



图 3　知识图谱应用架构图

Fig. 3　Diagram of knowledge graph application architecture

(2)功能应用：提供知识图谱场景化应用，负责完成资产多维画像、资产关联分析、资产分层画像以及漏洞资产查询四个功能的主要数据处理。

(3)数据平台：负责采集、整合、学习、推理和表示

图谱数据,提供知识抽取、知识融合、知识存储等服务。

4.2 应用场景

4.2.1 资产多维画像

对网络安全领域单个资产的性能参数指标、端口开放情况、地理位置信息、设备权属信息等进行整合,结合可视化技术,提供单个网络资产静态信息的全面描述。资产多维画像示意如图 4 所示。资产多维画像掌握的全量资产信息能够帮助资产管理人员和运维人员快速、直观地了解资产现状,为资产性能评估和状态监测提供便利条件,是图谱应用向外延伸的基石。



图 4 资产多维画像示意图

Fig. 4 Schematic diagram of multi-dimensional asset portrait

4.2.2 资产关联分析

资产关联分析通过挖掘资产间的内在联系与相互影响,借由资产间直接关系以及间接关系构建的资产关联网络,提供资产间的全方位视角,如图 5 所示。对于存在风险的单个资产,资产关联分析能够快速定位与其有关的所有资产,提升企业资产风险的防控能力;对于依赖关系脆弱、甚至闲置的资产,资产关联分析能够对其进行重点提示提醒,推动资产的合理配置与利用。

4.2.3 资产分层画像

资产分层画像在保持资产间原有关系的同时附加了层次关系,形成信息化资产硬件层、服务层、应用层三层画像。资产分层画像示意如图 6 所示,硬件层包含小型机、存储、交换机、网闸等;服务层包含操作系统、虚拟机、中间件等;应用层包含领域内的应用系统。

4.2.4 漏洞资产查询

漏洞资产查询示意如图 7 所示,漏洞资产查询将定期的网络安全漏洞扫描结果与信息资产知识图谱相结合,明确漏洞关联资产的 IP、操作系统、端口、运维负责人、业务负责人、应用系统等相关信息,并向相关负责人推送漏洞信息及整改方案,督促其及时高效处理。



图 5 资产关联分析示意图

Fig. 5 Schematic diagram of asset correlation analysis

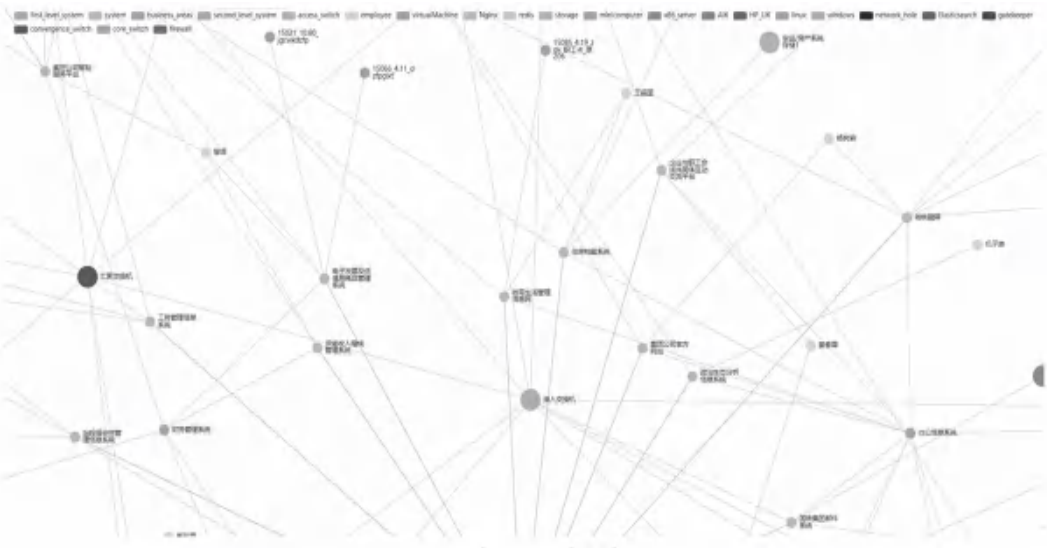


图 6　资产分层画像示意图

Fig. 6　Schematic diagram of asset stratification portrait



图 7　漏洞资产查询示意图

Fig. 7　Schematic diagram of vulnerability asset query

5　结束语

结合铁路信息资产自动化、精细化、智能化管理需求，研究构建了网络安全领域信息资产知识图谱，并基于该知识图谱设计了资产多维画像、资产关联分析、资产分层画像和漏洞资产查询四个应用场景，为路局集团公司挖掘资产潜在价值、准确定位漏洞风险提供了重要的数据支撑。未来，将开展更深层次的资产自动发掘、网络风险预警、网络态势推演等方面的研究工作，探索建立铁路网络安全领域资产快速应急与主动防御信息平台，助力知识图谱技术成为构建安全、智能网络环境过程中不可或缺的核心技术。

参考文献

[1] 牛勇. 网络安全知识图谱构建的关键技术研究[D]. 成都：电子科技大学，2021.

[2] 李新琴，刘承亮，代明睿，等. 铁路运输设备故障知识图谱构建与应用研究[J]. 铁路计算机应用，2023，32(1)：20-24.

[3] Wu X H，Duan J W，Pan Y，et al. Medical knowledge graph：Data sources，construction，reasoning，and applications[J]. Big Data Mining and Analytics，2022，6(2)：201-217.

[4] Shen X W，Li X Y，Zhou Bin，et al. Dynamic knowledge modeling and fusion method for custom apparel production process based on knowledge graph [J]. Advanced Engineering Informatics，2023，55：101880.

[5] Zou Z W，Wang B，Li F，et al. Research on network security threat analysis method based on knowledge graph [C]// Proceedings of 2024 IEEE 7th Advanced Information Technology，

Electronic and Automation Control Conference. Piscataway, NJ: IEEE, 2024: 501-507.
[6] 涂文奇, 李柏岩, 刘晓强, 等. NL2SQL 融合知识图谱在设备运维数据检索中的应用[J]. 智能计算机与应用, 2024, 14(9): 118-124.
[7] 田煜. 面向应急物资管理的知识图谱技术的研究与实现[D]. 北京: 北京邮电大学, 2022.
[8] 王宏, 朱海庭. 基于知识图谱的数字文物平台的研究与实现[J]. 智能计算机与应用, 2024, 14(10): 170-175.
[9] Yin L J, Zhang D Y. The calculation method of the network security probability of the multi-rail division based on fuzzy inference[J]. Mobile Networks and Applications, 2022, 27: 1368-1377.
[10] Neema H, Wang L Q, Koutsoukos X, et al. Model-based risk analysis approach for network vulnerability and security of the critical railway infrastructure[C]// Proceedings of the 16th International Conference on Critical Information Infrastructures Security. Cham: Springer, 2021: 79-98.
[11] 张丁荣, 王恪铭, 冯心妍. 基于知识图谱和故障树的高速铁路事故致因分析[J]. 铁路计算机应用, 2023, 32(7): 14-18.
[12] 王丹. 多源知识图谱事件知识融合方法研究[J]. 智能计算机与应用, 2024, 14(5): 157-163.
[13] Bonner S, Barrett I, Ye C, et al. A review of biomedical datasets relating to drug discovery: A knowledge graph perspective[J]. Briefings in Bioinformatics, 2022, 23(6): bbac404.
[14] 李琪, 孙宇宁. 铁路信息化资产知识图谱测绘工作的构想[J]. 铁路计算机应用, 2020, 29(8): 15-18.
[15] 钟晓峰, 杨国正, 单连勇. 网络安全知识图谱研究综述[J]. 信息对抗技术, 2024, 3(5): 19-25.
[16] Procházka J, Novobilsky P, Procházkova D. Cybersecurity of railway network management and partitioning[J]. Railway Report, 2020, 189: 125-130.
[17] 刘军, 冷芳玲, 吴旺旺, 等. 基于多模态和知识蒸馏的教材知识图谱构建方法[J]. 计算机科学与探索, 2024, 18(11): 2901-2911.
[18] 李志飞, 赵月, 张龑. 基于表示学习的知识图谱推理研究综述[J]. 计算机科学, 2023, 50(3): 94-110.
[19] Luo Y, Ruggiano N, Bolt D, et al. Community asset mapping in public health: A review of applications and approaches[J]. Social Work in Public Health, 2023, 38(3): 4-7.
[20] 王威龙. 基于知识图谱的资产管理系统原型设计与研发[D]. 杭州: 杭州电子科技大学, 2022.
[21] 卫婧, 向君, 朱广劼, 等. 铁路网络安全领域知识图谱构建及应用的研究[J]. 铁路计算机应用, 2022,31(11): 55-60.
[22] 姚奕, 杨帆. 联合知识图谱和预训练模型的中文关键词抽取方法[J]. 计算机科学, 2022, 49(10): 243-251.