网络空间正面临一个复杂多变的时代,一方面云、大、物、移、智等新兴技术深刻改变传统产业面貌,推动新兴产业崛起,发展的需求从未如此强烈;另一方面,网络病毒、恶意程序、固有漏洞、网络入侵行为等在网络空间大行其道,网络安全防护的需求从未如此强烈。新时期网络安全形势发生深刻变化,现实中针对工业互联网、物联网、关键信息基础设施的威胁与日俱增,时时掣肘经济社会发展,威胁国家安全和人民群众切身利益。事实证明,安全与发展相伴相生,网络安全防护作为"一体两翼"的支撑作用必不可少。十九届四中全会提出推进国家治理体系和治理能力现代化,亟需信息化、网络化技术为方方面面赋能,加强各个领域网络安全防护已经成为时代课题,开展这方面的研究也因此极具现实意义和学术价值。

——本期特约组稿人:数学工程与先进计算国家重点实验室教授　单征

一种基于知识图谱的工业互联网安全漏洞研究方法

陶耀东[1,2],贾新桐[2,3],吴云坤[3]

(1.北京交通大学,北京 100044;2.工业控制系统安全国家地方联合工程实验室,北京 100015;
3.奇安信科技集团股份有限公司,北京 100015)

摘　要:工业互联网安全问题日益突出,对工业互联网安全漏洞知识库的深入研究是解决问题的关键。为解决漏洞数据利用价值低、关联分析手段欠缺、可视化程度不足等问题,以工业互联网安全漏洞库为基础,提出了构建工业互联网安全漏洞知识图谱的方法,通过原始数据信息提取、关联关系分析、数据存储等手段,将知识图谱导入到 Neo4j 图数据库,以实现高效存储、查询。从时间维度、空间维度、关联关系维度进行知识图谱的分析,将查询结果进行了可视化展现。结果表明:提出的方法可以有效、直观地展现工业互联网安全漏洞数据的自身属性与关联关系,实现漏洞数据内在价值的深度挖掘。

关键词:工业互联网;漏洞库;知识图谱;关联关系

中图分类号:TP393　　**文献标识码**:A　　DOI:10.19358/j.issn.2096-5133.2020.01.002

引用格式:陶耀东,贾新桐,吴云坤.一种基于知识图谱的工业互联网安全漏洞研究方法[J].信息技术与网络安全,2020,39(1):6-13,18.

A research method of industrial Internet security vulnerabilities based on knowledge graph

Tao Yaodong[1,2],Jia Xintong[2,3],Wu Yunkun[3]

(1.Beijing Jiaotong University,Beijing 100044,China;
2.Industrial Control System Security National Local Joint Engineering Laboratory,Beijing 100015,China;
3.Qi An Xin Technology Group Co.,Ltd.,Beijing 100015,China)

Abstract:The problem of industrial Internet security is becoming more and more prominent. The key to solve the problem is to deeply study the knowledge base of industrial internet security vulnerabilities. In order to solve the problems of low value of exploiting vulnerable data,insufficient means of association analysis and insufficient degree of visualization,based on the data of industrial Internet security vulnerability database,a method of constructing knowledge graph of industrial Internet security vulnerabilities is put forward. Knowledge graph is imported into Neo4j graph database by means of original data information extraction,association analysis and data storage,so as to achieve efficient storage and query. From the time dimension,spatial dimension and correlation dimension,the knowledge graph is analyzed,and the query results are visualized. The results show that the proposed method can effectively and intuitively show the attributes and relationship of vulnerability data in industrial Internet,and realize the deep mining of the intrinsic value of vulnerability data.

Key words: industrial Internet;vulnerability database;knowledge graph;correlation

0 引言

随着信息化与工业化深度融合,工业互联网产业与相关技术获得了快速发展。信息化技术逐渐成为促进传统工业生产模式进行创新与发展的重要驱动力与技术保障。工业互联网产业蓬勃发展的同时,也面临着巨大的安全挑战。

一方面,工业互联网安全漏洞数量急剧增加。根据公共漏洞和暴露(Common Vulnerabilities and Exposures,CVE)等权威漏洞数据库的数据,本文提取出了各年披露的工业互联网安全漏洞,结果表明:自 1999 年工业控制系统安全漏洞首次被披露以来,漏洞数量总体呈增长态势,且 2010 年"震网事件"发生后,漏洞数量增速急剧增加。2010 年全年披露的工业互联网安全漏洞数量为 77 条,而 2018 年全年披露的工业互联网安全漏洞数量多达 637 条。在信息技术(Information Technology,IT)与操作技术(Operational Technology,OT)日益结合的今天,原本相对封闭的工业环境打开了一个个缺口,降低了攻击者的攻击门槛。对于攻击者来说,不仅学习成本和攻击难度下降,而且其可利用的漏洞数量显著增多,使工业互联网安全面临着严峻的挑战。

另一方面,传统工业控制系统在工业互联网大环境下的暴露面巨大[1]。工业控制系统内的设备、主机、工业协议等普遍存在版本和型号老旧的问题。在实际生产中,为了维持生产环境稳定性,往往不会对已存在的漏洞进行修复、对软硬件进行升级更新。一个典型的案例是永恒之蓝事件,由于工业主机系统版本老旧,且没有及时打补丁,工业生产环境成为"永恒之蓝"事件的重灾区。

为了应对工业互联网面临的日益严峻的安全挑战,加强工业互联网安全漏洞知识库的建设是必由之路。但是目前该领域的工作起步较晚,而且研究手段缺乏进一步的创新,导致工业互联网安全漏洞知识库不能充分发挥其在安全防护工作中的重要作用,难以体现其蕴含的丰富价值。本文提出了一种利用知识图谱相关技术手段对工业互联网安全漏洞进行研究与分析的方法。研究了工业互联网安全漏洞信息从结构化数据转储为知识图谱的方法,并使用 Neo4j 图数据库实现知识图谱的高效存储、查询。为了深度挖掘漏洞信息价值,设计并实现了信息提取引擎以及关联分析引擎,不仅能够从时间维度、空间维度、漏洞各字段维度进行分析,而且能够从漏洞、事件、产品的关联关系维度进行分析并实现可视化展示。结果表明:该方法可以深度挖掘工业互联网安全漏洞在安全特性、时间特性、空间特性、关联关系等维度的价值,并且能够多维度、全方位地对工业互联网安全漏洞知识图谱进行可视化展现。

1 数据来源与相关技术

1.1 工业互联网安全漏洞库

工业互联网安全漏洞库(Industrial Internet Security Vulnerability Database,ISVD)包含了自 1999 年以来记录的 3 727 条工业互联网安全漏洞,是本文进行研究与分析的基础数据信息。该漏洞库既包含传统工业控制系统安全漏洞,又包含新时代网络安全环境下工业互联网领域涉及的安全漏洞,具有数据来源权威、信息渠道多样、统计字段丰富、信息准确性高、安全事件关联等特点。

(1)权威性

ISVD 的漏洞数据主要来源于公共漏洞和暴露(Common Vulnerabilities and Exposures,CVE)数据库、中国国家信息安全漏洞共享平台(China National Vulnerability Database,CNVD)、中国国家信息安全漏洞库(China National Vulnerability Database of Information Security,CNNVD)、美国国家漏洞库(National Vulnerability Database,NVD)、中国关键基础设施安全应急响应中心、美国工业控制系统网络应急响应小组等。以上列出的公开漏洞信息发布来源都是信息安全领域内最具权威性的平台。

(2)多样性

一方面,ISVD 收集漏洞信息的渠道具备多样性。既包括公开漏洞信息发布平台发布的漏洞,又包括安全研究团队自身挖掘出的漏洞,还收录了其他白帽子提供的漏洞信息。漏洞信息渠道的多样性保证了 ISVD 能够汇集各个平台、安全团队的优势,形成合力。

另一方面,ISVD 内漏洞信息维度多样。由于漏洞信息有不同的来源,对于同一条漏洞实体,不同来源能够提供不同的漏洞信息维度。例如,CVE 虽然在"漏洞描述"方面做得十分细致,但是并不会提及该漏洞的危害评分。而 CNNVD 和 NVD 会在漏洞评分方面给予数据支持。目前,ISVD 中有 24 个维度,可以全面地描述漏洞信息。

(3)准确性

为了进一步核实与确保漏洞信息的准确性,ISVD 中记录的漏洞信息会由工业自动化专家与信息安全专家共同审核与验证。工业互联网安全漏洞的特殊性在于,研究者必须既熟练掌握工业控制系统领域的知识,又深刻了解互联网领域信息安全技术的攻防技术。因此,多方专家联合工作是保障漏洞信息准确性的必要条件。

(4)关联性

传统的信息安全漏洞库大多没有将漏洞信息与实际发生的安全事件结合起来,而只是关注于漏洞信息本身。但在长期的工业控制系统安全攻防实战中可以清楚地看到,绝大多数由漏洞引起的安全事件都具备相当高的相似性。这就意味着了解已经发生的安全事件,是指导工业互联网安全防护工作的重要一环。ISVD 将安全事件的攻击手段、利用漏洞、影响范围、防范手段等做了详细的记录,有助于相关单位与企业进行借鉴,防患于未然。

1.2 知识图谱

知识图谱是一种结构化的、具备强大语义处理及关联分析能力的语义知识库[2-8]。知识图谱的概念是谷歌公司于 2012 年首次提出的[9]。随着互联网应用的普及和人工智能技术的发展与应用,该领域在短短数年内获得了长足的发展,在学术界和工业界都诞生了许多研究成果[10]。知识图谱可分类为通用知识图谱与领域知识图谱。通用知识图谱不聚焦于某一专业领域的数据信息,更关注知识的广泛性,并以此为基础开发出丰富的上层应用。不同的是,领域知识图谱更关注于研究者的专业领域内的知识,侧重于知识的深度,对于其他领域专业的研究者来说易读性较低且应用门槛较高。

知识图谱的价值在于它能够内生地在海量数据中挖掘数据间的深层次关系,让数据之间的关联关系产生创造更多的价值。Freebase、DBpedia 和 Zhishi. me 是通用知识图谱中的典型代表。其中,Freebase 是谷歌旗下的开放共享知识库,允许任何人创建、修改、查询。目前 Freebase 的业务已经迁移至 WikiData[11],但其通用知识图谱的实质并没有发生变化。DBpedia[12]是摘取维基百科的信息形成结构化数据,是世界上最大的多领域知识本体之一。DBpedia 强化了维基百科的功能,并将其他资料连接至维基百科。Zhishi. me[13]通过从开放的百科数据中抽取结构化数据,首次尝试构建中文通用知识图谱。

知识图谱常用的描述方式包括 RDF(Resource Description Framework)、RDFS(Resource Description Framework Schema)和 OWL(Web Ontology Language),上述方法可以对实体或资源进行准确的描述[14-15]。知识图谱的数据源往往有以下两类:一是网页数据,该类数据源广泛用于通用知识图谱的构建;二是数据库,该类数据源是领域知识图谱的主要数据基础。对于数据库中存储的结构化数据来说,构建知识图谱的工作主要在于结构化数据到领域本体的映射以形成知识图谱,并将知识图谱进行存储。

知识图谱的存储方式分为 RDF 存储和图数据库存储两类。其中,RDF 存储是指:将知识图谱数据以 S-P-O(Subject,Predicate,Object)三元组的形式进行存储。虽然 RDF 存储的查询、归并、链接效率较高,但是其更新和维护比较困难。图数据库存储相比较来说更适合进行知识图谱的查询与存储,而且支持各类图挖掘算法。Neo4j 是使用最为广泛且十分成熟的图数据库[16-18]。Neo4j 支持 Cypher 查询语言,可以灵活地实现节点和关系的表示与查询等操作。本文使用 Neo4j 存储工业互联网安全漏洞知识图谱,并使用 Cypher 进行增、删、改、查等操作。

2 工业互联网安全漏洞知识图谱构建

本文提出的基于知识图谱的工业互联网安全漏洞分析方法如图 1 所示。首先从 ISVD 漏洞库原始数据中提取出全部的 3 727 条工业互联网安全漏洞信息,形成漏洞节点。此外,对漏洞原始信息进行半自动化分析,提取出漏洞信息中涉及的工业互联网安全事件以及工业互联网产品,分别形成事件节点和产品节点。在提取出各类节点之后,将节点信息注入关联分析引擎,对事件与漏洞、漏洞与产品、产品与事件这三类关联关系进行梳理与记录。至此,工业互联网安全漏洞知识图谱关键组件:节点、节点间的关联关系都已经生成。将节点与节点间关联关系存储至 Neo4j 图数据库的方式有两种:一是利用 Python 库 py2neo 实现数据的插入;二是使用 Neo4j 官方提供的工具 Neo4j-import,将节点信息文件 nodes. csv 与关联关系信息文件 relation. csv 导入 Neo4j。

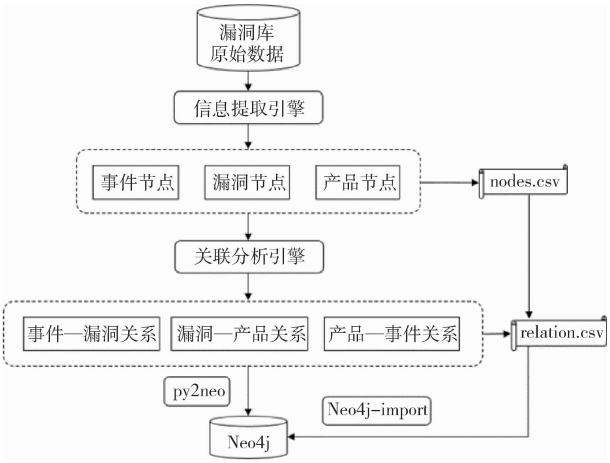

图 1　工业互联网安全漏洞知识图谱分析方法

2.1　原始数据信息提取

ISVD 中的漏洞信息是本文进行研究与分析的基础数据信息。主要包括如下字段:isvdId(ISVD 漏洞 ID)、name(漏洞名称)、dateTime(发布时间)、cveId(CVE-ID)、riskLevel(风险级别)、vulType(漏洞类型)、description(漏洞描述)、deviceList(影响设备)、refer(参考资料)、supplier(供应商)、patch(补丁信息)、score(漏洞评分)、eventId(事件 ID)、eventDescription(事件描述)等。每一个漏洞信息条目不仅对漏洞自身属性进行了详细的描述,而且对漏洞涉及的产品和事件进行了记录。但是产品和事件信息与漏洞信息杂糅在一起,难以梳理出漏洞、产品、事件之间的逻辑关系。因此,本文的工作在此基础上进行了信息的提取与关联分析。

信息提取引擎的工作,目的是将漏洞信息、事件信息、产品信息从原始数据中提取出来,并形成节点,用于知识图谱的构建。

首先,介绍信息提取引擎的提取规则。

规则 1:漏洞节点中的 eventId、eventDescription 字段,直接映射到事件节点中的 eventId、eventDescription。

规则 2:通过正则表达式在 eventDescription 中匹配事件时间的描述,若存在,则规范化为 * * * * - * * - * * 的形式并记录在 eventTime 字段;若不存在,则将漏洞节点的 dataTime 直接映射到事件节点的 eventTime。

规则 3:将漏洞节点中的 deviceList 直接映射到事件节点的 deviceList,若同一 eventId 对应多条漏洞,则将 deviceList 进行增量补充。

规则 4:将漏洞节点中的 cveId、cnnvdId、cnnvdId 等唯一标识漏洞条目的字段,按照上述优先级直接映射到事件节点的 vulnerabilities,若同一 eventId 对应多条漏洞,则将 vulnerabilities 进行增量补充。

规则 5:遍历漏洞节点,将其中的 deviceList 字段内容进行分割,并赋予唯一性标识 productId:PID,规范化数据以自动化方式直接映射,非规范化数据使用人工处理的方式。

规则 6:将漏洞节点中 deviceList 切割后的内容直接映射到产品节点的 productVersion 字段。

规则 7:若漏洞节点中的漏洞描述字段(description)包含漏洞涉及产品的相关描述(模糊匹配),则将匹配到的内容映射到产品节点的 productDescription。

规则 8:将漏洞节点中的 cveId、cnnvdId、cnnvdId 等唯一标识漏洞条目的字段,按照上述优先级直接映射到产品节点的 vulnerabilities,若同一产品对应多条漏洞,则将 vulnerabilities 进行增量补充。

其次,介绍各类节点的构成以及节点示例。对于漏洞节点来说,其内部构成与 ISVD 漏洞数据结构和内容没有大的差别,较为完整与详细地记录了漏洞的自身属性以及相关的产品与事件属性。以 ISVD-2014-0824 漏洞为例,其漏洞节点部分信息如表 1 所示。

表 1　ISVD-2014-0824 漏洞节点

isvdId:VID	name	cveId	…	eventId	eventDescription
ISVD-2014-0824	Microsoft Windows OLE 远程执行代码漏洞	CVE-2014-4114	…	ISED-2014-0001	事件使用的攻击样本 0day CVE-2014-4114 此 0day 影响……

对于事件节点,维度相较于漏洞节点少一些。一方面,安全事件自身的敏感性导致其事件细节不会被广泛公开,这就给信息安全研究者的事件分析带来了难度;另一方面,安全事件数据侧重于语言描述,缺乏结构性信息。事件 ID(eventId)用来唯一标识事件,事件描述(eventDescription)从事件发生

背景、攻击手段、事件影响等方面进行说明,事件发生时间(eventTime)记录事件发生的时间属性,设备列表(deviceList)记录了此次安全事件中主要涉及的存在漏洞的产品信息,漏洞(vulnerabilities)记录了该事件涉及的漏洞编号。以事件 ISED-2014-0001 为例,其事件节点信息如表 2 所示。

表 2 ISED-2014-0001 事件节点

eventId:EID	eventDescription	eventTime	deviceList	vulnerabilities
ISED-2014-0001	事件使用的攻击样本 0day CVE-2014-4114 此 0day 影响……	2014-10-15	Microsoft Windows Vista SP2, Windows Server 2008 SP2……	CVE-2014-4114, CVE-2014-0751

产品节点记录了工业互联网安全漏洞涉及的主要产品的 ID、产品版本、产品描述及其相关漏洞编号。由于原始数据中对产品的描述并不多,因此当前产品节点信息量不够丰富。在后续的工作中,建设工业互联网产品知识库并与 ISVD 进行信息共享以实现优势互补,是一个重要的研究方向。以产品 ISPD-2018-0001 为例,其产品节点信息如表 3 所示。

表 3 ISPD-2018-0001 事件节点

productId:PID	productVersion	productDescription	vulnerabilities
ISPD-2018-0001	U. motion Builder <1.3.4	U. motion Builder 是国外某公司的一款生成器产品	CVE-2014-4114, CVE-2014-0751

2.2 关联分析

知识图谱中的另一个重要构成要素就是关系(Relationship)。关系,将各个分散的节点关联在一起,表现出逻辑性,体现了数据的深层次价值。工业互联网安全漏洞知识图谱关注于三类关系:事件—漏洞关系、漏洞—产品关系、产品—事件关系。关联分析引擎的作用就是以前文所述节点为基础,对这三类关联关系进行挖掘并存储。

首先,介绍关联分析引擎的关联规则。

规则 9:遍历事件节点,对于 eventId:EID、vulnerabilities 字段数值同时存在的情况,对每一个从 eventId:EID 到 vulnerabilities 的映射,建立一个从事件 ID 指向漏洞 ID 的关系。

规则 10:遍历产品节点,对于 productId:PID、vulnerabilities 字段数值同时存在的情况,对每一个从 productId:PID 到 vulnerabilities 的映射,建立一个从产品 ID 指向漏洞 ID 的关系。

规则 11:遍历事件节点,对于 eventId:EID、vulnerabilities 字段数值同时存在的情况,以 vulnerabilities 为线索遍历产品节点,对每一个从 eventId:EID 到 productId:PID 的映射,建立一个从事件 ID 指向产品 ID 的关系。

其次,介绍各类关系的构成以及关系示例。

事件—漏洞关系包括事件 ID(:EID)、类型(type)、漏洞 ID(:VID)三个字段。以 ISED-2014-0001 事件为例,该事件涉及两个漏洞,类型字段为 include,两个漏洞分别是 ISVD-2014-0824、ISVD-2014-0015,如表 4 所示。

表 4 ISED-2014-0001 事件—漏洞关系

:EID	type	:VID
ISED-2014-0001	include	ISVD-2014-0824
ISED-2014-0001	include	ISVD-2014-0015

漏洞—产品关系包括产品 ID(:PID)、类型(type)、漏洞 ID(:VID)三个字段,如表 5 所示。

表 5 PID-2011-0396 产品—漏洞关系

:PID	type	:VID
PID-2011-0396	contain	ISVD-2018-0396
PID-2011-0396	contain	ISVD-2017-0385
PID-2011-0396	contain	ISVD-2017-0386
PID-2011-0396	contain	ISVD-2016-0108
PID-2011-0396	contain	ISVD-2015-0073
PID-2011-0396	contain	ISVD-2015-0138
PID-2011-0396	contain	ISVD-2011-0195
PID-2011-0396	contain	ISVD-2011-0199
PID-2011-0396	contain	ISVD-2011-0218
PID-2011-0396	contain	ISVD-2011-0238
PID-2011-0396	contain	ISVD-2011-0239

事件—产品关系包括事件 ID(:EID)、类型(type)、产品 ID(:PID)三个字段,ISED-2014-0001 事件为例,该事件涉及 7 个产品,类型字段为 involve,如表 6 所示。

表 6 ISED-2014-0001 事件—产品关系

:EID	type	:PID
ISED-2014-0001	involve	PID-2010-0231
ISED-2014-0001	involve	PID-2010-0233
ISED-2014-0001	involve	PID-2010-0235
ISED-2014-0001	involve	PID-2011-0067
ISED-2014-0001	involve	PID-2011-0068
ISED-2014-0001	involve	PID-2011-0069
ISED-2014-0001	involve	PID-2011-0070

2.3 数据存储

以上工作实现了工业互联网安全漏洞知识图谱中节点和关系的构建,已经形成了逻辑关系上的知识图谱,还需要将信息进行存储,才能实现对知识图谱的实际利用。本文选取了 Neo4j 图数据库来存储工业互联网安全漏洞知识图谱。该图数据库存储、查询十分高效,技术手段成熟,具备非常高的稳定性,同时还有丰富的接口,易于实现功能的扩展。

本文使用两种方式实现数据的存储:一是 py2neo 库,二是 Neo4j-import。

py2neo 是一个使用 Python 语言进行 Neo4j 数据库操作的库。由于信息提取引擎和关联分析引擎都是由 Python 语言实现的,因此使用 py2neo 来实现存储具有连续性,而且便于在存储过程中实现数据的进一步处理。主要存储方法为:

(1)创建一个 Graph 类的实例 graph;

(2)对于每一个节点,添加一个 Node 类的实例并添加相关属性,使用 graph.create()方法在图中创建该节点;

(3)对于每一个关系,添加一个 Relationship 类的实例并添加相关属性,使用 graph.create()方法在图中创建该关系。

Neo4j-import 是官方提供的知识图谱数据导入工具,具备占用资源少、导入速度快等特点。该方法的操作对象是.csv 文件。在执行导入操作时,必须先让 Neo4j 停止运行。并且,该方法的导入操作只能生成新的数据库,而不能在已有数据库的基础上增量添加数据。因此,本文在第一次数据导入时使用了 Neo4j-import 方法,在后续的操作中更多地使用 py2neo。

3 工业互联网安全漏洞知识图谱分析

构建后的工业互联网安全漏洞知识图谱存储在 Neo4j 图数据库中,通过 Neo4j 提供的 Web 管理页面,使用 Cypher 语言实现增、删、改、查等操作。本节从漏洞自身属性、关联关系角度来分析工业互联网安全漏洞知识图谱,将漏洞的深层次价值信息进行可视化的展示。

3.1 漏洞属性分析

漏洞属性反映了漏洞自身的基础标识信息、风险特征、补丁信息等。利用知识图谱对漏洞属性进行多维度分析,研究者不仅可以帮助更加深入地了解漏洞自身特征,也能够通过统计数据及其可视化展示结果更加清晰地把握工业互联网安全漏洞整体趋势。

从时间维度出发,对 2018 年工业互联网安全漏洞的危害等级进行分析。知识图谱中同时存在高、中、低、暂无四类漏洞危害等级。由于部分漏洞尚未披露出足够的细节,难以对其进行评分,因此定级为暂无,不进行统计查询。查询结果如图 2 所示。2018 年工业互联网安全漏洞中,高危漏洞数量为 343 个,中危漏洞数量为 232 个,低危漏洞数量为 20 个。2018 年工业互联网安全漏洞中高危漏洞数量占比约为 57.65%,占比较高。相比于文字描述,可视化手段在结果展现方面更加直观。

从空间角度出发,对 2018 年工业互联网安全漏洞中归属于我国的漏洞数量与其他国家漏洞数量进行对比。查询结果如图 3 所示。结果表明:2018 年,归属于我国的工业互联网安全漏洞数量为 16 条,归属于其他国家的漏洞数量为 415 条。虽然我国的漏洞数量占比较小,但是该结果也体现了我国在工业互联网领域的产品覆盖率较低。

3.2 关联关系分析

从关联关系的角度对漏洞、事件、产品进行分析,将知识图谱中节点间的“关系”进行可视化展现,深度挖掘工业互联网漏洞信息的价值。

事件—漏洞关系记录了工业互联网安全事件中涉及到的漏洞信息。以某断电事件为例,查询结果如图 4 所示,可以查询到该事件相关的 2 条工业互联网安全漏洞,分别是 CVE-2014-4114、CVE-2014-0751。除此之外,漏洞节点与事件节点可以进

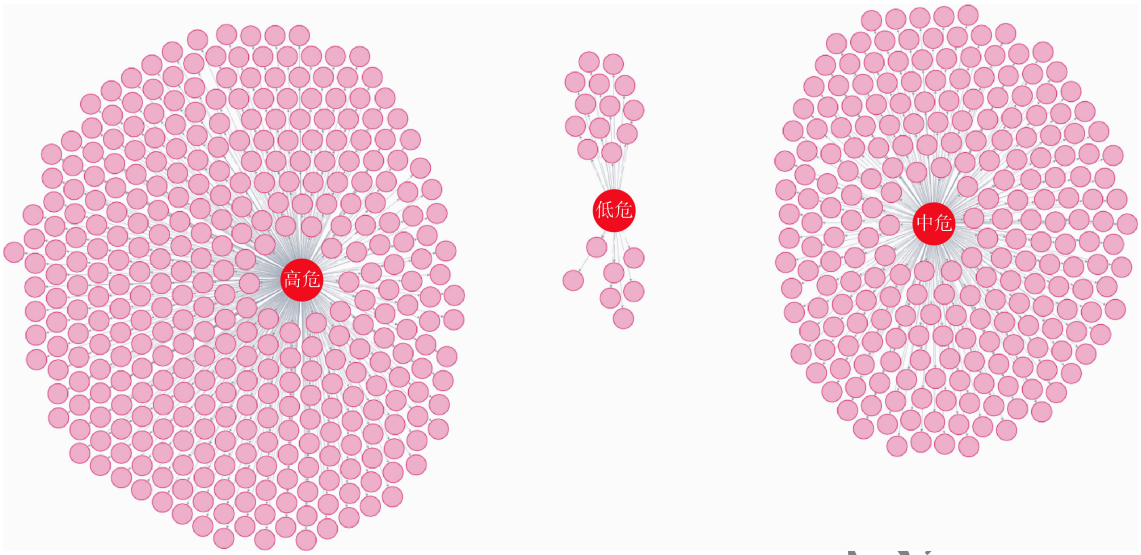

图 2　2018 年工业互联网安全漏洞危害等级

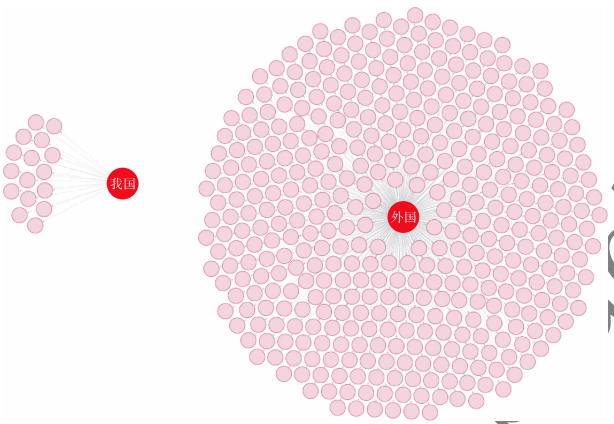

图 3　2018 年工业互联网安全漏洞空间分布

一步展开节点细节,供安全研究人员进行事件分析,进而用以指导类似工业互联网场景下的安全防护工作。

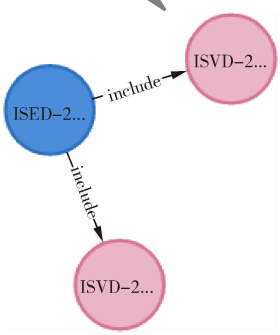

图 4　某断电事件—漏洞关联关系

产品—漏洞关系记录了工业互联网相关产品存在的已披露的安全漏洞信息。以某系列 PLC 为例,查询结果如图 5 所示。

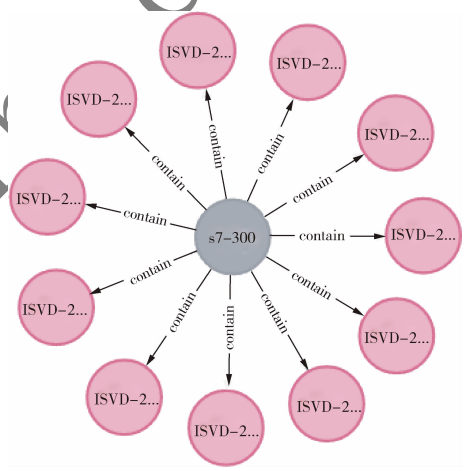

图 5　某系列 PLC—漏洞关联关系

事件—产品关系记录了工业互联网安全事件中涉及的产品信息。同样以某断电事件为例,查询结果如图 6 所示。结果表明:产品是安全漏洞的根源,是攻击者实施破坏的切入点。梳理工业互联网安全事件与相关产品的关系,有助于从资产管理、变更管理等角度进一步完善安全防护策略,提升整体安全防护水平。

4　结论

为解决工业互联网安全漏洞数据利用价值低、关联分析手段欠缺、可视化程度不足等问题,本文以工业互联网安全漏洞库数据为基础,提出了构建工业互联网安全漏洞知识图谱的方法,通过原始数据信息提取、关联关系分析、数据存储等手段,将知识图谱导入到 Neo4j 图数据库,以实现高效存储、查

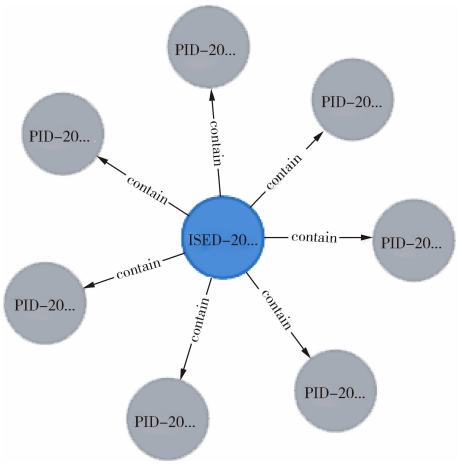

图 6　某断电事件—产品关联关系

询,并从时间维度、空间维度、关联关系维度进行知识图谱的分析,将查询结果进行了可视化展现,深度挖掘了工业互联网安全漏洞数据的内在价值。

后续的工作应从以下两点展开:一方面,原始漏洞数据的内容需要更加丰富,形式应当更加规范;另一方面,应加强漏洞、事件、产品、概念验证(Proof of Concept,PoC)代码、补丁信息的记录与分析。

首先,原始漏洞数据的内容需要更加丰富,形式应当更加规范。例如,在本文进行信息提取引擎的开发时,产品节点的提取遇到了困境:原始漏洞数据中产品相关内容匮乏,表述形式不规范、不统一。为了工作的顺利推进,采用了半自动化的方式进行分类、提取,对规范且全面的数据内容使用自动化手段,对形式不规范或者内容不完善的数据内容进行人工干预。因此,在后续的工作中,应丰富漏洞数据的维度,并进一步规范数据形式。该做法不仅有助于工业互联网安全漏洞库自身的完善,而且也提升了以此为基础的知识图谱的价值。

其次,应加强漏洞、事件、产品、概念验证代码、补丁信息的记录与分析。漏洞不是单一的、孤立的实体,而是包含于事件、存在于产品、通过 PoC 被利用、因补丁而修复的高价值数据。而其价值的挖掘与事件库、产品库、PoC 库、补丁库等知识库的建设息息相关。因此,在后续的工作中,应逐步建立漏洞库相关的一系列知识库,并进行多知识库的关联分析,掌握工业互联网安全漏洞的相关规律与特征,提升工业互联网安全防护整体水平。

参考文献

[1] 尚文利,安攀峰,万明,等. 工业控制系统入侵检测技术的研究及发展综述[J]. 计算机应用研究,2017,34(2):328-333.

[2] 贾焰,亓玉璐,尚怀军,等. 一种构建网络安全知识图谱的实用方法[J]. Engineering,2018,4(1):117-133.

[3] ASAMOAH C,TAO L,GAI K,et al. Powering filtration process of cyber security ecosystem using knowledge graph[C]. IEEE International Conference on Cyber Security & Cloud Computing. IEEE,2016.

[4] LI A. A practical approach to constructing a knowledge graph for cybersecurity[J]. Engineering, 2018, 4(1): 53-60.

[5] CAI Z, FENG Y, LIU S, et al. A new intrusion prevention model using planning knowledge graph[C]. International Conference on Graphic & Image Processing. International Society for Optics and Photonics,2013.

[6] MING J, JAVAD L, JOHANSSON K H. Power grid AC-based state estimation: vulnerability analysis against cyber attacks[J]. IEEE Transactions on Automatic Control,2018: 1-1.

[7] CHOPADE P,BIKDASH M. Structural and functional vulnerability analysis for survivability of Smart Grid and SCADA network under severe emergencies and WMD attacks[C]. IEEE International Conference on Technologies for Homeland Security. IEEE,2014.

[8] PATSOS D,MITROPOULOS S,DOULIGERIS C. Expanding topological vulnerability analysis to intrusion detection through the incident response intelligence system[J]. Information Management & Computer Security,2010,18(4): 291-309.

[9] SINGHAL A. Introducing the knowledge graph: Things, not strings[Z]. Official Google Blog,2012:1-8.

[10] AUER S,BIZER C,KOBILAROV G,et al. DBpedia: a nucleus for a web of open data[C]. The Semantic Web: ISWC'07/ASWC'07 Proceedings of the 6th International the Semantic Web and 2nd Asian Conference on Asian Semantic Web,2007:722-735.

[11] SUCHANEK F M,KASNECI G,WEIKUM G. Yago: a core of semantic knowledge[C]. Proceedings of the 16th International Conference on World Wide Web,2007:697-706.

[12] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]. Proceedings of 2008 ACM SIGMOD International Conference on Management of Data, 2008:1247-1250.

(下转第 18 页)

生成得不够精确,导致发现的路径不能被执行,降低了路径覆盖率。

本文提出了基于参数约束的分支覆盖符号执行优化算法,将具有特殊性质的参数进行收集并添加到约束集合中,使得生成的测试用例更加精确,可以执行到具有特殊参数的路径中,提高了路径覆盖率。目前只对函数的特殊参数进行了约束,下一步的研究工作是将实际程序中具有特殊性质的选项进行收集并利用,生成约束集。

参考文献

[1] 李舟军,张俊贤,廖湘科,等. 软件安全漏洞检测技术[J]. 计算机学报,2015,38(4):717-732.

[2] 谢肖飞,李晓红,陈翔,等. 基于符号执行与模糊测试的分支覆盖导向性混合测试方法[J/OL]. 软件学报:1-18[2019-11-18]. https://doi.org/10.13328/j.cnki.jos.005789.

[3] 叶志斌,严波. 符号执行研究综述[J]. 计算机科学,2018,45(S1):41-48.

[4] 汪孙律,林渝淇,杨秋松,等. 基于输入约束的符号执行优化[J]. 通信学报,2019,40(3):19-27.

[5] TRABISH D, MATTAVELLI A, RINETZKY N, et al. Chopped symbolic execution[C]. The 40th International Conference on Software Engineering. ACM,2018: 350-360.

[6] GODEFROID P. Compositional dynamic test generation[C]. Proceedings of the 34th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages,2007,42:47-54.

[7] RAMOS D A, ENGLER D R. Under-constrained symbolic execution:correctness checking for real code[C]. The 24th Usenix Conference on Security Symposium. USENIX Association,2015: 49-64.

[8] WONG E,ZHANG L,WANG S,et al. DASE:document-assisted symbolic execution for improving automated software testing[C]. 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering (ICSE). ACM,2015: 620-631.

[9] 郭曦,王盼. 基于依赖条件重构的程序符号值分析方法[J]. 电子学报,2019,47(3):630-635.

[10] 安靖,钟金鑫,魏更宇,等. 函数摘要在 Concolic 测试方法中的应用[J]. 北京邮电大学学报,2012,35(1):24-27.

(收稿日期:2019-11-21)

作者简介:

於家伟(1994 -),男,硕士研究生,主要研究方向:网络与信息安全、漏洞挖掘。

李世明(1976 -),通信作者,男,硕士,副教授,主要研究方向:网络空间安全物联网技术、数据挖掘技术。E-mail:hsdlsm@163.com。

毕雪洁(1994 -),女,硕士研究生,主要研究方向:网络与信息安全、漏洞挖掘。

(上接第 13 页)

[13] Niu Xing, Sun Xinruo, Wang Haofen, et al. Zhishi.meweaving Chinese linking open data[C]. The Semantic Web-ISWC,2011:205-220.

[14] 齐斌,王宇,邹红霞,等. 基于信息熵和模糊集的网络安全知识图谱选择技术[J]. 小型微型计算机系统,2018,39(11):48-52.

[15] CHAWUTHAI R,TAKEDA H. RDF Graph Visualization by Interpreting Linked Data as Knowledge[M]. Semantic Technology,2015.

[16] HOKSZA D,JELINEK J. Using Neo4j for mining protein graphs:a case study[C]. International Workshop on Database & Expert Systems Applications,2015.

[17] WEBBER J. A programmatic introduction to Neo4j[C]. Conference on Systems,2012.

[18] DRAKOPOULOS G,BAROUTIADI A,MEGALOOIKONOMOU V. Higher order graph centrality measures for Neo4j[C]. International Conference on Information,2016.

(收稿日期:2019-10-21)

作者简介:

陶耀东(1981 -),男,博士,研究员,主要研究方向:工业互联网安全、数控技术。

贾新桐(1994 -),男,硕士,主要研究方向:工业互联网安全。

吴云坤(1975 -),男,硕士,主要研究方向:大数据安全。

版权声明