面向网络空间安全情报的知识图谱综述

董 聪, 姜 波, 卢志刚, 刘宝旭, 李 宁, 马平川, 姜政伟, 刘俊荣

中国科学院信息工程研究所 北京 中国 100093
中国科学院大学网络空间安全学院 北京 中国 100049

摘要 随着网络空间安全情报在网络犯罪、网络战和网络反恐等领域的作用日益凸显, 迫切需要对网络空间安全情报的基本理论和综合分析方法进行深入研究。当前, 安全情报在实际应用中主要面临着数据类型多样、分布离散、内容不一致等问题, 因此引入知识图谱技术框架, 旨在利用知识图谱面向海量数据时信息收集及加工整合的思想, 提高安全情报的收集效率、情报质量, 同时拓展情报的使用范围。本文首先简要回顾安全情报和知识图谱的研究现状, 同时介绍知识图谱在安全领域的应用。其次给出面向安全情报的知识图谱构建框架。然后介绍安全情报知识图谱构建的关键技术, 包括信息抽取、本体构建和知识推理等。最后, 对安全情报知识图谱发展面临的问题进行了讨论。

关键词 网络空间安全; 安全情报; 知识图谱; 信息抽取; 本体构建; 知识推理
中图法分类号 TP391.1 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2020.09.05

Knowledge Graph for Cyberspace Security Intelligence: A Survey

DONG Cong, JIANG Bo, LU Zhigang, LIU Baoxu, LI Ning, MA Pingchuan, JIANG Zhengwei, LIU Junrong

Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract With the increasingly prominent application and role of cyberspace security intelligence in cybercrime, cyberwarfare, and network counter-terrorism, it's urgent to intensive study the basic theories and effective extraction methods of cyberspace security intelligence. At present, security intelligence mainly faces the problems of diverse data types, discrete distribution, and inconsistent content. Therefore, the knowledge graph technology framework is introduced to improve the security intelligence, which aims at using the knowledge graph to solve the problem of the information collection and processing integration of massive data, improving the collection efficiency and intelligence quality of security intelligence. This paper first briefly reviews the research status of security intelligence and knowledge graph, and shows the application cases of the knowledge graph of security intelligence in intelligence analysis. Second, it summarizes the framework for building a knowledge graph of security intelligence. Then, it introduces the key technologies for the construction of security knowledge graph, including information extraction, ontology construction and knowledge reasoning. Finally, the issues facing the development of security intelligence knowledge maps are discussed.

Key words cyber security; security intelligence; knowledge graph; information extraction; ontology construction; knowledge inference

1 引言

随着信息化的不断扩大以及网络技术的持续发展, 网络攻击的方式也在逐渐成熟, 呈现出长持续性和高隐蔽性的特点, 尤其是在大国博弈中, 高级可持续威胁(Advanced Persistent Threat, APT)[1]成为国家间网络对抗的主要手段。传统的防御手段如入侵检测(Intrusion Detection System, IDS), 入侵防御(Intrusion Prevention System, IPS)等在面对大规模的结构化应用系统时略显不足, 因此兴起了网络空间安全情报[2-4], 网络空间安全态势感知[5]等综合防御策略。

通讯作者: 姜波, 博士, 副研究员, Email: jiangbo@iie.ac.cn。
本论文获得中国科学院网络测评技术重点实验室和网络安全防护技术北京市重点实验室, 国家自然科学青年基金(No.61702508, No.61802404)、北京市科委重大研究项目(No.D181100000618003)、中国科学院战略性先导 C 类(No.XDC02000000)资助。
收稿日期: 2018-08-02; 修改日期: 2018-11-04; 定稿日期: 2020-08-24

安全情报通过信息的共享, 减少安全风险, 增强整体的安全性[6]。安全情报包括漏洞情报、威胁情报、资产情报等可用于安全分析的相关信息。其中, 近几年兴起的威胁情报(Threat Intelligence)是一种基于证据的信息集合, 用于描述针对资产的威胁信息, 例如恶意 IP 地址, 恶意样本描述, 攻击者特征等。与其他类型情报相比, 威胁情报更侧重于己方系统之外, 攻击者及攻击工具的信息描述[7]。使用威胁情报分析敌手的攻击行为, 可以了解到自身系统的不足, 并随之做出相应的调整, 将被动防御变为主动防御。因此, 威胁情报研究引起了学术界和工业界的广泛关注, 当前已有针对威胁情报共享机制[8-11], 威胁关联分析[12-14]等研究方向的探索, 同时威胁情报共享平台也逐步投入实用, 如 X-Force-Exchange, ThreatBook, 360TI, Virustotal, Threatcrowd 等。

安全情报提高了网络主动防御的能力, 但是在其发展和应用方面仍面临诸多问题。首先, 从海量数据中提取高价值的安全情报存在一定难度。当前安全情报主要通过人工提交收录的方式增加数量, 缺少从开放网络信息主动生成安全情报的能力。其次, 安全情报尤其是威胁情报的离散性分布严重。不同安全情报库中存在信息关联程度较低, 甚至相互冲突的情况, 降低了安全情报的可信性。另外, 安全情报的综合使用率较低。威胁情报、漏洞情报等发展较为独立, 缺少与资产情报的融合分析, 难以充分发挥安全情报整体的威力。在信息检索领域中广受关注的知识图谱技术的兴起, 为解决当前安全情报研究面临的问题提供了一种整体的思路。

知识图谱(Knowledge Graph, KG)[15]并非一个完全崭新的概念, 其原型是 1998 年由 Tim Berners-Lee 提出的语义网(Semantic Web)[16]。语义网的初衷是使用语义链接代替无语义链接将互联网信息连接起来, 但是由于多方面的原因, 语义网的发展较为缓慢[17]。直到 2012 年, 谷歌借鉴语义网络技术, 提出知识图谱的概念, 并宣布用知识图谱技术来提升检索效果, 从而带动了知识图谱技术在信息检索领域的研究。迄今为止, 成熟的知识图谱产品不断投入到实际应用中, 如谷歌的 Knowledge Graph, 微软的 Satori, 搜狗的知立方等。同时存储通用知识的知识库也在逐渐完备, 如 Freebase[18], DBpedia[19]等。在信息检索领域的成功, 使得知识图谱技术受到越来越多的关注, 其他领域也相继利用这一技术辅助与支撑实际应用场景。例如, 在金融领域, 知识图谱技术被用于股票的分析[20]以及金融诈骗的推理[21]。在公安情报领域, 知识图谱技术被用于辅助线索分析, 预防电信诈骗[22]等。

知识图谱通过信息抽取、知识融合、知识推理等过程[23-24], 将分散在多处以不同形式表示的信息进行关联融合, 形成一个统一表示且高质量的知识集, 继而根据现有的知识进行推理, 挖掘潜在的知识同时产生新的知识, 从而实现安全情报分析的智能化。基于知识图谱对信息的整合能力, 安全情报知识图谱将在如下实际场景中发挥作用: (1)安全情报搜索。在情报库中查找相关情报是较为常见的应用, 准确查找到不同类型的情报将减轻情报分析的工作量。知识图谱将搜索视为实体的搜索而非简单的字符串搜索的思想[15], 可用于构建知识层级的查询系统, 达到提升情报查询结果的相关程度及查询效率的目的; (2)敌手画像构建。画像构建是根据用户或团体的属性信息构建用户模型的常用方法。基于威胁情报等来源对敌手的常用工具、攻击手法、社工情报等信息进行收集关联, 知识图谱可以构建详细描述敌手信息的画像, 展示攻击者的全貌, 更精准的实现攻击溯源; (3)团伙情报挖掘。网络攻击行为通常由多人或多个团伙发起, 但在要素众多的情报中挖掘团伙信息面临着困难。知识图谱从主体、事件、人和物等语义层面构建情报的关联关系, 并根据设定的规则进行挖掘从中寻找线索, 可实现团伙情报分析以及隐匿组织的发现; (4)APT 攻击发现: APT 攻击是当前互联网领域面临的严重威胁, 具备 APT 攻击的检测能力是实现网络安全的重要保证。当前, 通过单一的数据分析实现 APT 检测的概率较低, 需要探索多维度联合的分析方法。知识图谱可以将资产、威胁、漏洞、流量、日志等信息进行统一描述, 打破数据鸿沟, 并进一步应用知识推理的方法实现异常行为的分析, 从而实现 APT 的发现。

目前, 针对安全情报知识图谱的研究和应用仍较少, 因此, 本文首先通过对知识图谱现有通用技术以及在网络安全领域的应用进行调研总结, 归纳出面向安全情报的知识图谱构建框架。然后, 对其关键技术进行系统梳理, 旨在将知识图谱技术引入安全情报领域。最后, 探讨知识图谱技术在安全情报研究与应用中仍需解决的问题。

本文的组织结构如下: 第 2 节简要介绍安全情报的发展和通用知识图谱构建技术, 给出安全情报知识图谱的应用场景; 第 3 节提出安全情报知识图谱构建框架; 第 4 节梳理安全情报知识图谱构建的关键技术, 并讨论存在的问题; 第 5 节展望安全情报知识图谱的未来研究方向。在下文中, 若无特殊说明, 使用情报知识图谱指代安全情报知识图谱, 情报知

识代替安全情报知识。

2 背景知识

2.1 安全情报研究

安全情报用于提升安全分析能力, 主要包括漏洞情报、资产情报以及威胁情报。不同的安全情报具有不同的内容, 漏洞情报关注于软件、硬件或协议的脆弱性带来的安全威胁, 资产情报包括企业或公司内部的软硬件资产以及对资产重要程度的描述信息[25], 威胁情报主要收集与攻击者或攻击行为相关的外部要素, 聚焦于收集、整合、共享威胁信息, 提供安全分析查询和比对的依据, 从而达到对威胁信息的及时管控[26-27]。其中, 漏洞情报发展较为成熟, 以围绕 NVD、CNNVD 等几个大型漏洞库建设、共享为主。威胁情报发展较晚, 但在近几年发展迅速。2012 年, 美国开始着手建立覆盖其关键基础设施的威胁情报共享体系, 开始形成威胁情报中心的雏形。2015 年, 美国众议院通过了网络安全信息共享法案, 提高了威胁情报在不同部门间的流通、共享、利用的能力。2017 年, 中国国家发展改革委批复了国家网络空间威胁情报 CNTIC(China National cyber-space Threat Intelligence Collaboration)共享开放平台的建设方案, 通过政府和企业合作共建的方式加强威胁情报的整合利用。除受到国家层面的重视外, 安全厂商通过建立自己的威胁情报中心, 提出威胁情报的相关标准和格式, 促进威胁情报的商业化。2012 年, MITRE 提出了结构化威胁表示框架 STIX(Structured Threat Information eXpression)成为情报表示的主流标准, 随后在此基础上进行改进提出 STIX2.0。其他相关威胁情报标准也相继被提出, 如 TAXII(Trusted Automated eXchange of Indicator Information)提供安全情报交换的协议, OpenIOC 提供灵活的事件表示框架, CybOX(Cyber Observable eXpression)从主机、流量等底层角度对情报信息进行划分。与此同时, 学术界也展开对威胁情报的研究, 如情报的溯源分析, 质量评估, 威胁推演等[28-30], 推动了威胁情报的发展。

2.2 知识图谱研究

知识图谱是谷歌用于增强其搜索引擎功能的辅助知识库[31]。学术界对知识图谱定义为以语义连接为基础的把现实世界的实体或概念联系起来的知识库[18, 32-33]。根据应用领域的不同, 知识图谱可划分为通用知识图谱和领域知识图谱[24]。其中, 谷歌及其他用于检索领域的知识图谱属于通用知识图谱范畴, 而本文研究的情报知识图谱则属于领域知识图谱。

通用知识图谱的基本构成单元是知识三元组, 即(实体-关系-实体)的形式。通过实体和关系的相互连接构建网状的知识结构。通用知识图谱的构建大致分为抽取、融合、加工、评估和推理的过程[23-24]。通过抽取过程进行实体识别和关系识别得到信息素材, 然后实体对齐和关联合并实现知识融合, 此时的知识是无结构扁平化的知识, 进一步在得到的知识上进行聚类分析和本体构建实现层次化的知识梳理, 并通过质量评估和知识挖掘提高知识的质量, 最终实现知识的推理与实际使用。知识图谱的发展得益于多方面技术的提高, 深度学习和自然语言处理的发展提高了信息抽取的准确性和鲁棒性[34], 知识的嵌入式表示是面向知识图谱中的实体和关系进行表示学习, 在低维向量空间中高效计算实体和关系的语义联系, 为知识获取、知识融合和知识推理提供新思路[35]。此外, 最初用于语义网的 URI、OWL/RDF、SWRL 等技术标准也通常被应用于知识图谱的构建。

3 情报知识图谱构建框架

情报知识图谱旨在借助知识图谱技术对分散的安全情报进行整合, 实现情报聚合分析和应用场景扩展等目的。情报知识的来源包括安全分析报告、博客、社交网络、漏洞库、威胁情报库等, 构成要素包含而不局限于流量、样本、漏洞、域名、地址、主机、用户、组织、资产、攻击策略、攻击手法等。与通用知识图谱相比, 本文研究的安全情报知识图谱具有以下特点: 首先, 数据特点不同。与知识图谱相比, 情报知识图谱的覆盖范围有限, 仅关注特定领域的数据, 在数据规模以及要素规模上均小于通用知识图谱。同时, 情报知识图谱面向的信息具有专业特征, 例如信息的表示具有一定的特点, IP 地址、域名、漏洞等以固定的格式表示。其次, 知识与应用结合紧密。通用知识图谱的构建以知识的广度为主, 首要目标是构建涵盖各范围的知识以供智能搜索场景使用, 而情报知识图谱除了对大范围知识的覆盖外, 还需实现深度知识体系的构建, 达到知识体系与业务应用相适应的目的。例如, 在使用情报知识图谱分析捕获样本时, 在获取样本行为、攻击目标、编译路径等信息后, 不仅需要实现相关主体的查询, 而且期望能够应用于推断受害范围、分析使用漏洞、关联攻击组织等较为具体的业务应用。因此, 情报知识图谱的构建与通用知识图谱的构建不尽相同, 尤其体现在信息抽取、本体构建、知识推理与应用等过程中。本文将知识图谱构建技术与安全情报知识

的特点结合, 借鉴通用知识图谱的构建框架对情报知识图谱构建进行归纳, 如图 1 所示。与通用知识图谱构建框架相同, 情报知识图谱的构建过程同样包括三个层次: (1)信息抽取, 包括实体抽取、关系抽取和属性抽取; (2)知识融合, 实现多源异质信息的形式层面与内容层面的融合, 包括实体链接、本体工程、质量评估的过程; (3)知识加工与应用, 主要实现知识的后端处理, 包括知识存储、知识表示和知识推理。以下对情报知识图谱构建过程进行详细阐述。

从广泛的数据源中获取信息, 是构建情报知识图谱的首要环节。随着 Web2.0 的发展, 互联网中的信息量呈爆炸式增长, 安全情报数据的增长也不例外。在安全情报发展初始阶段, 手工识别信息的速度尚可与信息增长的速度相匹配, 但是随着近年情报研究的深入, 情报数据的增长速度加快, 以手工方式获取情报信息将耗费大量的人工与时间成本。因此实现信息的自动化获取, 是安全情报知识图谱构建的重要基础。信息获取的目标是得到实体和关系及属性信息, 同时需要根据数据源的结构化程度选择合适的抽取方法[36-37]。结构化信息的价值密度较高, 仅需少量的处理便可完成信息抽取[36]; 非结构化信息的价值密度较低, 需要通过复杂的处理过程, 其中以规则提取和统计模型抽取为主[38]。虽然结构化程度高的信息易于提取, 但是从另一个角度看, 结构化程度越高的信息其时效性越低。这是由于结构化信息是经过第三方对非结构化信息的加工而得到, 从而损失了信息的时效性。为了兼具提取的便利性和信息的时效性, 信息获取需要具备从不同结构化程度的数据源中提炼信息的能力。在 4.1 节中将总结现有的安全信息抽取方法。

得到实体, 关系及其上下文之后, 需要经过实体链接的过程消除歧义。实体链接的目标是解决不同数据集间数据的表示格式、指代内容存在差异的问题, 包括实体消歧和共指消解[24]。以下面的句子为例:

1. 通信模块是云计算平台中的必要模块。

2. 木马利用通信模块与 C&C 服务器通信。

3. 经调查发现, Control and Command 服务器地址为 xxx。

其中, 1 与 2 中的通信模块指代不同主体的通信模块, 属于需进行实体消歧的情形。2 中的 C&C 服务器与 3 中的 Control and Command 服务器同指命令和控制服务器, 属于需进行共指消解的情形。这两个过程的实现可充分借鉴通用知识图谱的相关研究方法[39-41]。

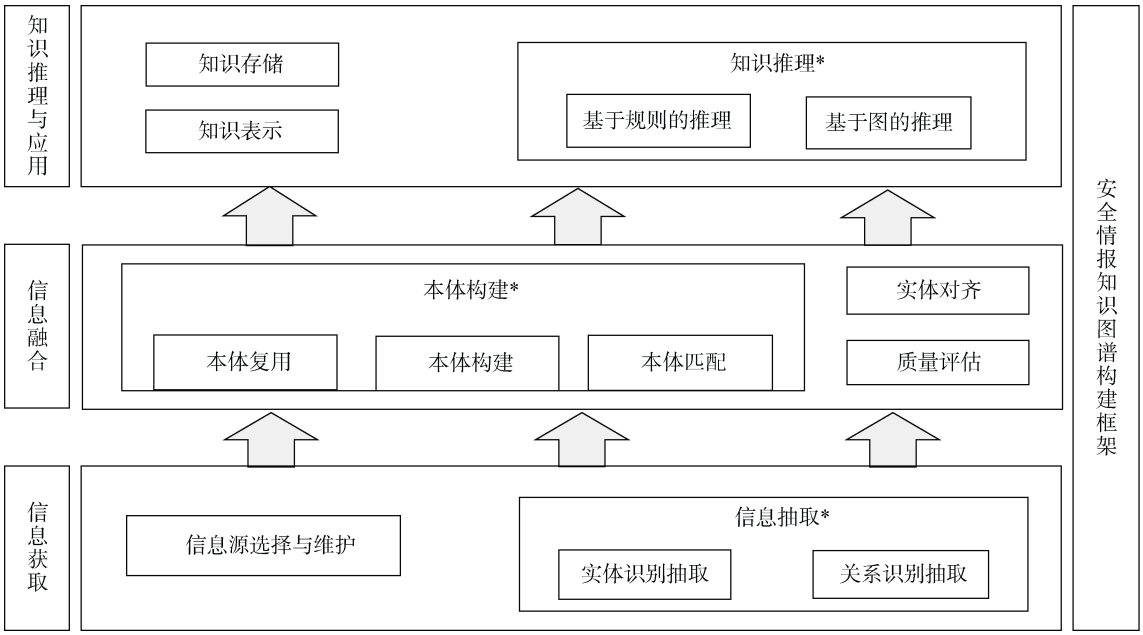

图 1　情报知识图谱构建框架

Figure 1　The Framework of Intelligence Knowledge Graph Construction

经过实体链接后, 信息仍停留在扁平化的结构上, 知识的相互联系较为单一且不充分, 因此需要通过本体工程, 完成知识融合的过程。本体是一种形式化的用于对共享概念体系明确而又详细的说明[42], 通过对具体知识的分类聚合实现知识的组织, 以及定义在本体上的关系和公理进行推理实现知识的延展, 因此本体构建是知识融合过程中的关键。本体构建有人工编辑和数据驱动两种构建方式[24]。在初始

构建本体的过程中, 由于情报知识图谱的数据规模较小, 以数据驱动的方式构建本体将面临数据源不足的问题, 同时难以构建现有知识的完整体系[43]。因此以人工编辑的方式手动构建初始安全情报的本体, 在规模上具有可行性, 同时可以更高效的还原现有知识体系[44-45]。本体构建并非一次性完成, 随着技术的发展, 知识体系也会发生变化并反映到数据中, 因此本体也需要通过更新过程与数据保持一致。在本体更新过程中采取以数据驱动自动构建本体为主并辅以人工审核的方式, 将更有利于新知识的补充。本体构建作为知识图谱构建的中心环节, 不仅实现知识语义层面的信息融合, 而且为后续的质量评估、知识推理与应用结合等提供语义依据[46]。在 4.2 节中将对现有安全本体进行梳理。

质量评估对信息融合后的知识进行质量校验, 避免质量低的知识存入到知识库中。情报知识图谱的质量评估主要关注: (1)信息相关度。由于安全情报数据与非安全数据混杂存在, 造成了严重的信息抽取噪声, 因此需要判断生成的情报是否属于安全情报, 或衡量与安全情报的相关程度; (2)冗余信息。多个数据源中可能存在同样的信息, 提前检测情报知识图谱中是否已存在同样的知识, 避免相同知识多次存入情报知识图谱中; (3)冲突信息。多个信息源中可能会存在互为冲突的信息时, 需进行真值判断发现真实的知识, 进而对冲突信息丢弃或者加入标记后再存入知识库中。

经过一系列处理后的知识需要保存到知识库中, 由于情报图谱中存在大量的关系型信息, 使用结构化数据库进行存储将产生大量的冗余存储信息, 因此将图数据库作为知识图谱的存储容器成为流行的选择[47]。当前较为常用的图数据库主要有 Neo4j 等。

存储于情报图谱中的信息通过知识推理过程实现知识的丰富以及与应用的结合[24]。目前, 情报数据组织形式较为简单, 数据间的关系难以展现, 主要以手工方式根据信息特征对信息建立关联, 这种方式在数据量巨大、数据源沟通不充分的情况下会面临效率低的问题。知识图谱通过赋予情报之间语义联系, 通过公理和规则在现有知识的基础上进行缺失知识的补全、隐含知识的挖掘以及与现实数据的结合, 从而实现情报内容的自动分析推理, 达到对现有信息的充分利用[48]。由此可见, 知识推理对于情报研究而言是构建知识图谱的重要一环。不同的推理方法涉及不同的知识表示, 传统的知识表示以三元组表示为主, 即(实体, 关系, 实体)的集合。W3C 公布的 RDF[49]为三元组表示提供了标准化形式。三元组的表示形式具有直观的特点, 但是在推理应用上不够高效。近几年兴起的知识分布式表示通过嵌入式方法将实体及其关系信息表示为低维向量, 简化了知识推理的计算, 受到了广泛的关注[35, 50], 例如 Trans 系列算法[51-54]。在 4.3 节将对现有的情报知识图谱推理方法进行总结。

4 情报知识图谱构建关键技术

情报知识图谱构建由信息抽取、本体构建、知识推理等关键过程组成。信息抽取实现实体、关系与原始数据的分离从而得到知识单元; 本体构建将碎片化的知识联系起来构建出知识网络; 知识推理通过现有知识产生新的知识, 在知识网络化的基础上进一步丰富知识, 同时与应用结合发挥价值。根据情报数据的特点以及安全业务流程, 情报知识图谱构建将从安全领域中已有的相关研究展开, 如: 日志、流量等信息的抽取, 安全资产本体、安全概念本体的本体构建, 基于规则和基于图形式的推理与应用等。本节通过对现有研究的整理分类, 在知识图谱框架下对这些技术重新进行审视, 总结适用于情报知识图谱构建的技术和方法要求。

4.1 情报信息抽取

情报信息抽取面向不同结构的数据从中自动抽取实体、属性及关系构成知识单元[55], 知识单元使用(实体, 关系, 实体) 三元组的形式表示。其中, 实体指安全活动中的主体信息, 例如漏洞、样本、病毒、事件等。关系是安全实体间相互联系的关系, 如攻击者与漏洞的关系, 病毒和恶意行为的关系等。属性信息则包括漏洞的发现日期、编号、描述、相关引用等。在一些研究中也将属性作为实体来看, 同样本文也将属性抽取划归到实体抽取中, 以减少抽取流程的复杂性[23-24]。得益于自然语言处理技术的发展, 自动化从海量异构的文本中抽取情报信息已有较多的研究成果可以借鉴, 主要可以分为两个思路: 基于规则匹配的方法和基于统计学习的方法。下文将从这两个角度对适用于安全情报抽取的研究进行归纳整理, 总结比较现有方法的优缺点及经验。

4.1.1 基于规则匹配的方法

基于规则匹配的方法通过对抽取过程多个步骤的分解, 利用预定义规则并结合机器学习算法实现信息的特征识别定位从而实现抽取[55-58]。基于规则匹配的方法具有准确、可靠、高效的特点, 在信息抽取与信息识别中使用广泛[59], 例如在入侵检测领域, Snort、l7-filter、Bro 等产品中的深度包检测技术也使用基于规则匹配的方法进行攻击类型的识别[60]。通

过与机器学习方法相结合构成多步骤的信息抽取方法, 减少规则的数量或自动生成规则, 解决匹配效率与抽取准确率平衡的问题, 是基于规则匹配方法的主要优势。

文献[61]提出正则表达式和本体相结合的方法抽取日志文件中的实体。该方法首先使用支持向量机判断日志文件是否与安全相关, 然后使用分隔符对格式相同的段落进行切分, 下一步通过遗传算法生成的正则表达式对段落中的信息进行标记, 最终通过本体匹配将标记信息转化为实体。该方法的优点是将半结构化文件中的格式作为特征用于类型识别以及生成正则表达式, 同时以信息抽取和本体匹配验证的方法提高抽取的准确率。但是该方法无法适用于非结构化文件的提取。

文献[62]提出正则表达式和语法树相似度结合的方法提取博客文本中的攻击指征(Indicator of Compromise, IOC)。该方法首先通过上下文词库和正则表达式对潜在的实体和关系进行定位, 然后对定位后的词进行语法树解析, 再与已有的标准语法树进行相似度计算构造特征矩阵, 最后将特征矩阵输入线性分类器判断是否为真正的实体及关系。利用安全特征词作为定位 IOC 指标的依据, 并且将安全实体抽取和安全关系抽取相结合是该方法的优点。

Bootstrapping 思想是适用于数据集中仅部分数据含有标签的半监督学习算法框架[63-64]。在信息抽取领域, 基于 Bootstrapping 的方法通过对基于规则方法的抽取流程的改进, 利用少量编写的规则即可自动生成大量规则。其中主要包括两个不断循环的过程: 一个是在文本数据中搜索已有的实体并根据其上下文模式产生规则加入到规则库中; 另一个是使用规则在文本数据中寻找符合规则的实体加入到实体库中。通过两个步骤的不断迭代, 可以逐步实现数据集中全部实体的标记[65]。利用 Bootstrapping 思想可以提高基于规则匹配方法的适用性及效率。

文献[66]提出了使用 Bootstrapping 方法用于从非结构化文本中提取安全关系。关系通常使用元组的形式表示, 例如 django hasVulnerability CVE-2017-7234, hasVulnerability 即是一种关系, 而(django, hasVulnerability, CVE-2017-7234) 则是对关系的完整表示。因此对关系的抽取可以转化为对元组的抽取[67]。该方法针对安全领域的关系定义了三种抽取模式: (1)两个实体类型间单个词的匹配; (2)两个实体类型间连续词集的匹配; (3)解析树的路径相似判断。同时为了提高 Bootstrapping 方法的准确率, 该方法使用了主动学习的方法和评分机制, 用于减少错误模式的生成。

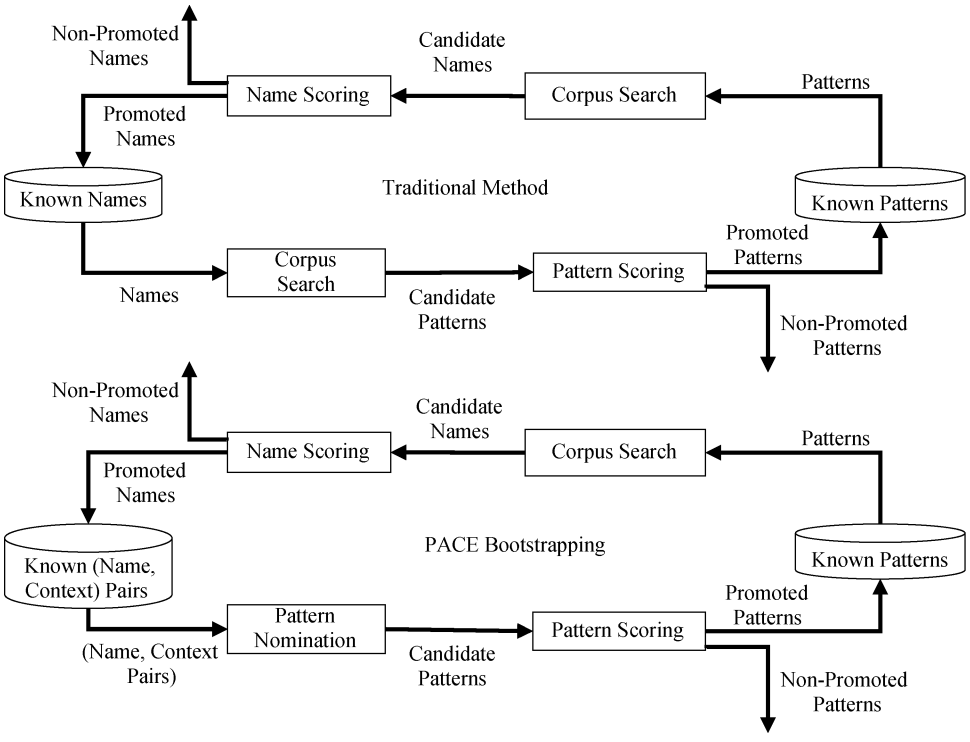

图 2 传统 Bootstrapping 方法与 PACE Bootstrapping 方法[68]

Figure 2 Traditional Bootstrapping Method and PACE Bootstrapping Method[68]

文献[68]提出了改进的 Bootstrapping 方法用于从博客、推特等文本中提取安全实体。传统的 Bootstrapping 方法一次循环需要两次全文搜索, 而 PACE 对此进行了改进, 使其在一次循环中只需要一次全文搜索。首先, 用一个带有上下文的实体库取代实体库和规则库(模式库)。在利用初始规则抽取实体时, 不仅对实体进行抽取, 同时选取其前后一定数量的词作为上下文词共同组成抽取结果。其次, 将规则的生成过程由在全文中进行搜索生成, 改为由在包含上下文词的实体库中生成, 从而减少了一次搜索全文的时间。除此之外, PACE 放宽了对生成规则的限制, 提高了对相关上下文词的选取要求, 从而在增加召回率的同时, 保持较高的准确率。

4.1.2 基于统计学习的方法

基于统计学习的方法利用最大熵[69]、条件随机场[70]、隐马尔可夫[71]等统计模型或词袋模型[72]进行语言关系的建模, 发现不同语言要素的统计规律, 实现实体与关系的识别。与基于规则的方法相比, 基于统计学习的信息抽取方法不需要人工构建规则, 而是自动从训练语料中学习参数, 较为简便。

随着机器学习的发展, 出现了较多的信息抽取工具, 例如斯坦福自然语言处理工具 Stanford NLP、自然语言处理工具包 NLTK、清华关键词抽取包 THUTag 等。然而这些工具并非针对安全领域所设计, 因此安全实体和安全关系的抽取效果并不理想。文献[73]和文献[74]使用 OpenCalais 对安全博客和论坛上的非结构化文本数据进行实体信息的提取。实验结果表明, OpenCalais 对于安全领域的实体识别精度不佳。在用于 IOC 的抽取中, Stanford 工具集的实体识别的准确率和召回率分别为 70%和 50%, 同时关系抽取的准确率和召回率分别为 50%~90%以及 10%~50%[75-76]。这主要由于不同的语料语言规律差别较大, 基于通用领域的语料训练的信息抽取模型不适用于特定专业领域的信息抽取。Stanford NLP、OpenCalais 等工具在 CoNLL 2003、MUC6、ACE2002 等通用领域的语料库上训练, 因而对于安全信息的识别和抽取效果较差[77]。

针对以上问题, 文献[78]和文献[79]在条件随机场模型上使用安全语料进行模型训练, 使得安全实体的抽取精度有所提升。条件随机场模型考虑了前后文词对当前词的影响, 除了对安全实体的识别较好外, 对于前后关联较紧密的函数名称和系统状态等名词同样具有较好的提取效果。但是上述实体抽取模型所需的安全语料大多采用人工提取的方式, 这种方式的缺点在于耗费大量人力和时间成本的同时仅产生规模较小的安全语料集。此外, 针对不同的应用场景, 仍需重新或部分标注模型所需的训练数据。因此, 如何获得足够的标注安全语料是阻碍该方法大规模应用的主要问题。

为了解决垂直语料的问题, 文献[80]提出了结合语料自动标记的安全实体抽取方法。该方法首先基于数据库匹配、启发式规则、安全词集三种方式对结构化文本如 NVD 漏洞库中的数据进行 IOB(Inside Outside Begin)标签的标注, 然后以 IOB 标签为特征构建最大熵模型实现非结构化文本中安全信息的提取。该方法从结构化文本中获取安全文本的训练语料, 为统计模型的训练问题提供了解决思路。使用统计模型可以根据语言和语义特征实现安全信息的抽取。但是对于垂直语料数量和质量的需求以及如何提高模型的抽取准确率, 仍是基于统计学习方法需要解决的主要问题。

4.1.3 总结与讨论

自然语言处理技术的发展促进了信息抽取研究的进步, 产生了大量文本处理方法, 其中基于规则匹配的方法以及基于统计学习的方法在安全信息抽取领域均有相应的研究。表 1 总结了安全信息抽取代表性工作及方法特点的对应关系。基于规则匹配的方法通过构建正则表达式或其他启发式规则实现信息的定位和提取, 并且与机器学习方法结合降低制定规则的人工消耗, 具有高效、准确的特点。但是基于规则匹配的方法在实现过程中较为复杂且不够灵活, 对新实体的识别存在困难。基于统计学习的方法使用训练语料构建统计学习模型如最大熵模型、条件随机场模型等, 可以实现自动化的信息抽取, 具有简便、鲁棒的特点, 适用于非结构化文本的抽取, 同时可实现新实体的识别。但是基于统计学习的方法存在抽取准确性低、严重依赖训练语料等问题。

数据是构成知识图谱的主体, 丰富的数据是成熟安全图谱的标志。从已有数据中准确的发现安全相关的实体和关系是知识图谱构成的关键。根据调研结果, 本文对适用于情报图谱的信息抽取经验总结如下: (1)抽取之前进行预处理, 判断是否属于安全类文本。由于信息抽取面向的是开放网络, 在海量文档中存在较多非安全领域的文档, 使用支持向量机等方法对文档进行过滤, 排除非安全文档, 从而减少抽取信息的时间, 同时可降低抽取的错误率; (2)融合基于规则与基于统计的方法实现信息抽取。安全领域中的实体和关系通常以安全术语的形式存在, 与通用知识图谱中的名词性实体和动词性关系不同,

安全实体和关系本身具有一定的特征, 例如漏洞的表示方法, 域名、地址的表示方法, 样本的哈希表示方法等均有一定的规律, 因此适用于启发式规则的查找。此外安全关系也可以通过上下文词实现语句定位后, 通过语法分析以及语义分析进行抽取。从另一个角度看, 用基于规则的方法抽取的信息虽然难以覆盖新出现的实体, 但可以基于现有数据扩充垂直语料, 为信息抽取精度的提升奠定基础。

表 1 情报信息抽取方法

Table 1 The Methods of Intelligence Information Extraction

类型	文献	抽取方法描述	性能评价			抽取对象		适用数据类型	
			准确率	召回率	F1	实体	关系	非结构化	半结构化
基于规则匹配的方法	[61]	正则表达式与本体结合抽取	0.828	0.782	0.80	√			√
	[62]	正则表达式与语法树解析、线性分类器结合抽取	0.98	0.92	NA	√	√	√	√
	[68]	基于 Bootstrapping 改进的 PACE 抽取方法	0.90	0.38	NA	√	√	√	√
	[66]	Bootstrapping、语法树解析、路径相似判断结合抽取	0.82	0.24	NA		√	√	√
基于统计学习的方法	[73]	SVM 文档相关性判断以及 OpenCalais 抽取	0.70	0.5~0.9	NA	√	√	√	√
	[74]	OpenCalais 抽取	0.50	0.1~0.5	NA	√	√	√	√
	[78]	条件随机场与安全本体结合抽取	0.83	0.76	0.80	√		√	√
	[80]	最大熵模型抽取	0.837	0.764	0.80	√		√	√
	[79]	条件随机场抽取	0.867	0.813	0.84	√		√	√

(注: NA 表示该方案无法参与该评分项; √表示该方案适用数据类型。)

4.2 情报本体构建

本体是同一领域内不同主体之间进行交流、连通的语义基础[81]。本体由多个元素组成, 其形式化定义如下:

$$\upsilon := (C, R, H^C, rel, A^\upsilon)^{[82]}$$

其中, C 是本体概念的集合, 通常使用自然语言进行描述; $H^C \subseteq C \times C$ 是上下文关系的集合, 定义了本体的层次结构; R 是非上下为关系, 其中的 $rel: R \to C \times C$ 定义了实际关系的映射; A^υ 是本体上公理的集合。其层次结构如下图所示。

安全情报本体作为情报知识图谱构建的核心层次, 是将信息抽取得到的实体及其关系构建为知识网络, 实现数据向知识的转化以及知识与应用的结合的过程。利用本体中定义的约束与规则可为后续的质量评估、知识推理等过程提供基础。概括的说, 本体的意义是专业领域知识与具体数据的结合。在安全领域, 本体的研究较为广泛, 但是尚无统一的安全本体可供借鉴[83], 当前的研究主要集中在安全的特定领域展开, 例如态势感知、入侵检测、漏洞挖掘、物联网安全等[84-87], 这些研究成果为情报知识图谱的构建提供了基础。当前, 本体主要以人工编辑的方式构建, 其原因是所涉及的数据类型较少, 使用人工编辑方式效率较高[24]。本文按照本体构建时面向知识模式和面向具体数据的两个角度出发, 将现有安全本体研究分为模式层本体以及数据层本体, 为安全本体的构建提供内容经验, 同时总结适用于情报知识图谱本体的构建方法。

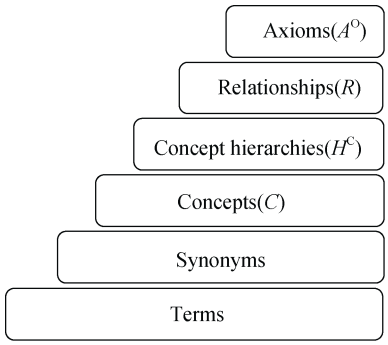

图 3 本体构建层次[82]

Figure 3 Layers of the Ontology Development Process[82]

4.2.1 基于模式的知识本体

模式层知识本体从网络安全研究的原理、需求、规范等抽象角度进行构建, 为确定知识范围, 构建知识框架、简化需求分析等提供支持。模式层知识本体实现抽象安全知识的梳理, 为数据层知识本体构建提供框架, 同时为信息抽取或知识推理等过程提供领域知识[88]等。

文献[89]针对研究语义网规范的通用本体框架(DOLCE-SPRAY)进行安全领域的拓展, 提出了安全本体框架 CRATELO。该框架包括三个层次, 分别为 DOLCE-SPRAY、SECCO(Security Core Ontology)以及 OSCO(Ontologies of Secure Cyber Operations), 囊括 223 个类别和 131 个关系。其中, SECCO 涵盖安全的主要内容, 包括安全需求、资产、威胁等内容, OSCO 涵盖安全性操作, 包括攻击性操作、防御性操作等分类。该本体在构建时利用层次化的思想, 提供了语义丰富、逻辑严谨的安全知识本体框架, 之后又有较多的研究对 CRATELO 进行丰富。

文献[90]提出 HUFO(Human Factors Ontology)对 OSCO 进行了丰富。HUFO 从人机交互的角度以及博弈的角度出发, 通过融合安全本体与信任本体, 将安全活动中人的能力、背景、动机和权限等因素转化为风险特征, 构建了人的内部特征和外部特征两个本体, 实现人为因素安全风险的评估和排序。但是为了对人的能力、背景、动机等因素进行定量分析和评价, HUFO 仍需进行更细粒度的扩展。

文献[91]使用业务流程和标记法构建安全需求本体。该本体通过符号化的方法对应用中的安全需求进行分析, 包括访问控制、隐私性、审计性、集成性、可用性、攻击检测/预防六个顶层需求本体以及三个层次的具体需求本体, 涵盖了较为全面的业务安全需求。但是此本体上的约束和规则较少, 缺少对需求的严格定义。

文献[92]描述了用于信息安全风险管理的本体。该本体基于文献[93]中的安全概念进一步的梳理, 描述了由威胁, 漏洞、控制、属性、评分组成的概念, 概念之间的关系以及基于 OWL 规范定义的形式化公理三部分构成安全领域的顶层知识框架。考虑到安全管理的具体应用, 该本体融合了安全标准 German IT-Grundschutz[94]作为顶层知识框架中更细粒度知识的填充, 使得该本体具有较强的实用价值。但是同样该本体缺少对概念的准确定义。

文献[95]描述了安全指标本体。该本体同样是一个层次化的本体, 通过对多个安全评价指标进行整合, 建立从多个维度度量整体安全性的本体, 包括五个顶层本体: 漏洞, 攻击, 态势, 防御和系统。其中系统本体处于核心位置, 系统的变化映射到两外四个本体的变化, 并通过指标来具体衡量变化程度。从效果来看, 该本体基于多个评价角度的构建, 可以更准确的对系统进行评价, 相对从分析角度和推理角度构建的本体衡量更为准确[83]。

文献[96]描述了用于攻击面定量分析的本体。由于攻击面分析只涉及资产状态的分析, 该本体并未使用分类完整但更为复杂的威胁情报共享框架, 例如 CyBOX 和 CIM(Common Information Model)等, 而是主要基于微软的 STRIDE 模型构建, 按攻击步骤构建了 6 个顶层本体, 包括系统, 攻击, 敌手, 防御, 任务和指标。基于该本体的攻击面推理系统给予了网络防御者在做出防御决策时平衡防御代价和系统安全的能力。

4.2.2 基于数据的知识本体

数据层知识本体从现有数据的格式、内容、结构化程度出发构建知识本体。根据面向数据层次的不同, 可以对数据层本体进一步分类。其中, 高层次数据本体面向语义关系丰富的情报数据, 对威胁实体、威胁关系、资产、漏洞等数据进行本体构建, 为情报融合、情报分析提供知识框架。在态势感知、入侵检测等研究中采用了通过本体建立状态联系的方法。这些研究通过本体对资产状态数据(日志、流量、系统记录等)进行关联匹配, 而后利用规则或其他方式实现数据融合。

文献[97]在 CRATELO 的基础上提出 OCNO (Ontology of Computer Network Operations)用于描述网络特征行为。该本体利用前后数据包短暂的连接关系将数据流转换为网络行为进行描述, 从而分析网络的状态。OCNO 是 CRATELO 中较为成熟的模块, 综合考虑了数据融合实现知识的生成以及知识的应用。 但是 OCNO 仅对传输层的行为进行了构建, 而未考虑对信息量更为丰富的应用层协议进行统一行为建模。

文献[61]提出了用于描述系统事件的本体。该本体包括对象本体(Objects Ontology)和事件本体(Event Ontology)。其中对象本体用于对主体信息的描述, 而事件本体将属于一个事件但分散在多处的对象本体关联起来, 进而从时间等维度进行集中分析和挖掘。该本体将结构不同的底层数据的信息进行统一表示, 实现信息的融合以及分析, 但是该本体的定义粒度较粗, 同时该本体上的约束与规则定义同样较少, 因此事件的分析仍需借助人工完成。

文献[74]提出用于融合多元异构数据的本体, 包括方法本体(Means Ontology)、结果本体(Consequences Ontology)和目标本体(Target Ontology)。该本体可对不同层次的数据如日志文件、流量数据、IDS 报警信息以及现有的威胁情报库进行融合。不同数据集中收集的信息将会对应到该本体中, 而后通过 OWL 断言的方式转化为三元组。通过预定义的规则进行一阶推理得到新的知识, 以达到准确发现入侵迹象的目的。

资产本体可以实现底层信息的实时整合, 但是相对而言只利用了内部信息。为了提高分析的准确率, 对资产本体进行扩充, 引入漏洞、威胁情报等外部信息, 可以提高信息融合的范围。文献[25]提出了用于威胁定量分析的本体。文献[25]以相关风险本体(Associated Risk)为中心, 将网络本体、STIX 框架、漏洞数据库 CVE 联系起来, 构建了扁平化的分析框架。通过建立在本体上的规则计算威胁最大似然相关性、识别受影响资产的过程, 实现情报信息的融合。同时利用提出的本体对 Red October 攻击事件进行分析, 验证了提出本体的有效性。而文献[98]则采用了层次化的思想, 以安全资产本体(Security Asset-Vulnerability Ontology)为中心向外扩展出威胁, 漏洞, 事件, 防御策略等多个本体, 较文献[25]的要素更为丰富。通过建立在本体上的规则将特征信息与漏洞、威胁等融合从而转化为风险评价。通过多个系统的部署, 在应对如 Mitnick 攻击[99]等针对分布式系统的攻击时具有良好的防御效果。

文献[100]构建了用于数据整合的 STUCCO 本体。基于安全知识图谱的开源项目, STUCCO 在威胁情报共享框架 STIX 以及分类标准 CybOX 的基础上, 考虑了与实体关系数据以及域名解析数据等不同层次数据的结合, 在尽可能保持简单和直观的基础上, 构建了一个包含漏洞、地址、个人等 15 个类别相关联的本体结构。STUCCO 从知识图谱的角度构建了面向威胁情报数据的安全本体, 但是 STUCCO 采用扁平化的构建方式, 而且约束规则较为简单, 因此其逻辑性、覆盖范围、可扩展性均较为欠缺。

虽然以上本体从各个角度出发提出了覆盖多种安全要素的本体, 但是各自较为独立, 并未考虑与其他本体标准的相互联系和整合, 对于构建全面的知识本体而言仍稍显不足。文献[101]提出了对当前威胁情报标准的整合和统一表示的本体 UCO(Unified Cybersecurity Ontology), 是当前较为全面实用的安全本体。虽然 STIX 在设计之初也考虑了对其他框架标准的融合问题, 但是 STIX 主要面向威胁情报信息, 而没有涵盖信息量较低的一些数据表示, 同时其 XML 格式表示信息不利于信息的自动推理。而 UCO 本体通过对现有威胁情报标准以及本体的研究, 通过相似类别合并和父类抽象的处理, 将多种标准融合为统一标准, 涵盖了当前主流标准表示的数据本体。UCO 采用 RDF/OWL 规范, 并且具有丰富的关系与约束规则, 因此可以支持信息的自动推理以及基于 SPARQL 语言的查询操作。

4.2.3 总结与讨论

安全本体的研究为情报知识图谱的构建提供了内容和方法的借鉴。表 2 总结了情报知识本体构建代表性工作及方法特点的对应关系。情报知识本体可以分为模式层本体和数据层本体: 基于模式的本体从抽象的知识角度出发, 关注于安全原理与安全需求, 为其他构建过程提供知识框架与需求分析; 基于数据的本体从数据的应用角度出发, 完成数据分类以及分析流程的构建, 实现知识与数据的结合。模式层本体与数据层本体互为补充, 共同构成完整的情报知识本体。其中, 模式层本体可以为数据层本体的构建提供领域知识, 形成数据层本体构建的理论支撑, 避免因涵盖数据不足导致的本体构建不完整; 数据层本体直接面向应用, 因此可以为模式层本体的构建提供分析范围, 明确安全需求以及安全问题, 避免本体构建与实际应用相脱节的情况。

表 2 情报知识本体构建方法

Table 2 The Methods of Intelligence Knowledge Ontology Construction

本体层次	文献	主要内容	优点	缺点
模式层	[89]	安全需求、资产、威胁	提出了可复用的框架	仍需进行完善
	[90]	人为因素、信任本体	将人的因素转化为特征	难以对具体指标定量评价
	[91]	安全需求	完整的需求框架	缺少与应用的结合
	[92]	资产、漏洞、威胁、风险管理标准	多层次架构, 充分利用已有知识	缺少概念的准确定义
	[95]	资产、漏洞、安全指标	利用已有的指标作为原子评价	本体数量较少, 缺少威胁情报的度量
	[96]	资产、威胁、攻击、防御	多层次架构, 本体定义规范	防御方式的建模不完整
数据层	[100]	漏洞、地址、组织、个人、软件、恶意代码	语义性丰富	扩展性欠缺
	[101]	威胁、脆弱性、漏洞、组织、事件、个人	融合多个本体, 支持推理和查询	实用性待检验
	[25]	漏洞、网络、STIX	量化威胁评价	本体要素较少
	[97]	网络协议	由框架扩展而来, 理论较完善	未包括高层次语义丰富的协议
	[61]	系统事件	从时间维度融合数据	规则约束较少
	[74]	流量、日志、告警	多源数据融合	可扩展性较差
	[98]	漏洞、资产、事件、防御策略	要素丰富、层次化	未融合现有威胁情报标准

综合两个层次不同本体构建的优点, 可总结适用于情报图谱本体的经验如下: (1)在现有设计的基础上进行本体设计。一方面, 即使仅针对网络安全领域, 使用手工编辑的方式从零开始构建本体也需要耗费大量的工作, 这是由于本体构建不仅要求对领域知识的精通, 而且要对本体设计流程和方法有所掌握。对于这两部分的要求使得仅有少数人具备从零开始构建安全本体的条件, 大多数人仍需补充较多的额外知识。另一方面, 当前大量的本体研究包含了安全概念、安全需求、安全分析、安全数据融合等不同范围层次的知识。从已有的研究出发, 对现有本体进行改良丰富, 直接采用或者一定程度上转化现有的本体, 可以减少构建安全知识体系的重复性工作, 同时避免一些设计缺陷及误区。(2)本体设计应具有层次性。一方面是由于知识本身具有层次特征, 因此使用层次化的本体设计可以与知识的自身组织相吻合。另一方面, 在不同的应用中, 知识的侧重不同, 需要从多个灵活的角度实现知识的融合, 因此使用层次化的本体有助于实现知识与数据、应用的结合, 从而充分发挥知识图谱的实用价值[102]。(3)本体设计应具有模块性, 即本体的构建应保持较好的可分性以及可扩展性。本体的构建不是一次完成的, 而是需要多次的迭代。基于当前知识构建的本体在数据量丰富的情况下会出现新的实体及关系。因此, 要尽量减少本体间的耦合性, 在力求全面的同时保持简单和直观, 为本体的扩充留有空间。(4)本体的设计应注重本体间关系以及约束规则的建立。情报知识图谱的本体并非仅是概念的划分, 更重要的是实现知识的关联融合, 使得知识孤岛通过建立在本体上的关系及约束规则, 形成相互关联与融合的丰富知识网络。

4.3 情报知识推理

推理是“使用理智从某些前提产生结论”的过程[103]。知识推理通过基于三元组(头实体, 关系, 尾实体)的知识表示, 利用已知三元组推测新的实体或关系要素组成新的三元组[24, 104], 实现新知识的发现, 对知识图谱进行补全。其过程如图 4 所示。

通用知识图谱推理主要研究如何提高推理的准确率, 而情报知识图谱推理则侧重于实现知识和业务的结合, 如何在风险评价、溯源取证、攻击路径推理、目标画像构建等实际安全应用场景中发挥其作用是当前情报知识图谱推理的重点研究方向及亟需突破的难题。因此, 情报知识图谱的推理在通用图谱推理的基础上仍需考虑如下方面: (1)知识与分析方法的结合。虽然对应到本体上的知识构成了相互连接的知识网络, 但是在场景应用中这些联系仍稍显不足, 通常需要借助于机器学习方法来更加有效的挖掘知识模式, 典型方法包括神经网络模型、马尔可夫模型、贝叶斯网络等。除此之外还可以在策略分析中引入钻石模型[99]和杀伤链(Kill-Chain)[106]模型等安全分析方法。(2)知识与低层次信息的结合。情报知识涵盖了高层次的历史信息, 但是像代码片段、流量包、数据日志等原始信息作为分析的对象并不包含在知识的范畴中。历史知识的内部推理具有一定的价值, 但是与现实信息的结合可以更充分的发挥安全情报的作用。(3)多元关系相互推理的结合。在现实中多元关系较二元关系更为普遍, 同样, 在情报推理中使用多元关系推理比二元关系更易于发现真实规律。

由于情报知识图谱与应用的紧密结合, 因此情报知识图谱的推理与通用知识图谱的推理具有不同的涵义, 其推理并不局限在三元组的推理中, 而是与不同的应用场景相适应, 具有多样的推理形式和推理结果。本文根据情报知识推理中借助的逻辑形式, 将现有安全研究的方法划分为基于规则的推理以及基于图的推理。

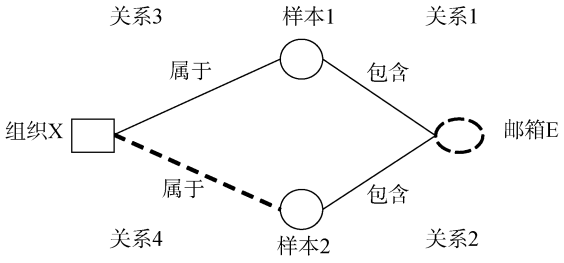

图 4 知识推理示意图.关系 4 可由关系 1、关系 2、关系 3 推出

Figure 4 Schematic of knowledge reasoning. Relation 4 can be inferenced by relation 1, 2 and 3

4.3.1 基于规则的情报知识推理

基于规则的推理方法借助规则、公理等逻辑形式实现知识的演绎推理。在安全研究中, 基于规则的推理方法主要包括基于一阶逻辑规则的方法和基于本体规则的方法[48]。其中, 基于一阶逻辑的方法通过构建谓词逻辑公式实现知识的推断, 具有较为久远的研究历史, 但是谓词逻辑公式的使用范围较窄, 且应用过程较为复杂。基于本体规则的方法在近几年的研究中较为充分。以 OWL 和 SWRL 或其他形式化语言定义的规则约束在本体的基础上建立推理关系, 具有定义简洁、描述丰富的特点[107]。

文献[108]将 SQL 作为规则使用, 用于判断

Android 中的访问策略是否存在不当配置。该方法的核心思想是利用访问模式将具体的访问行为与显式定义的访问策略联系起来, 用于鉴别未遵循最小化原则的访问策略, 如图 5 所示。该方法将预定义的访问策略存储到数据库中, 同时将测试用例中的访问行为抽象成访问模式, 最后使用自动生成的 SQL 检索是否存在与访问模式对应的访问策略, 从而实现逻辑的判断, 可以推断出定义不规范的访问策略。将 SQL 作为规则的载体, 可以在构建完整的逻辑模型后实现简单的真值判断, 但是无法承载复杂的语义, 其推理能力有限。SWRL 是用于语义网的推理语言, 在基于本体 OWL 表示的基础上, 可以实现丰富的逻辑推理。

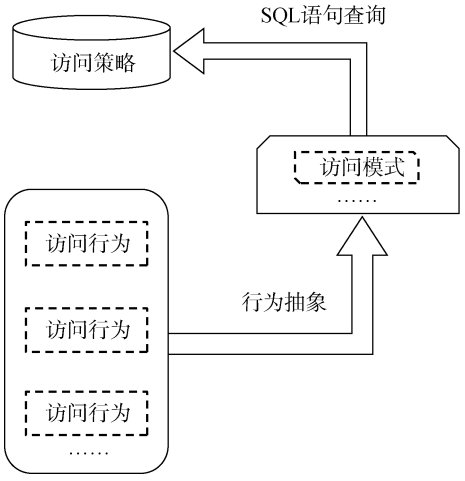

图 5　文献[108]的建模过程.

Figure 5　The modeling process of [108].

文献[109]基于 UCO 本体, 提出了使用 SWRL 规则对 Twitter 中出现的安全内容与内部资产情报结合产生针对性告警的方法。该方法将从 Twitter 中抽取安全内容匹配到本体上, 根据本体上的 SWRL 规则针对特定的系统画像产生告警, 可用于自动发现社交媒体上的安全信息。但是该方法中的规则的定义过于扁平化, 不利于规则的管理。为了实现大量规则的有效管理, 可以从时间维度和空间维度等多个角度定义规则, 提高规则的管理效率。

文献[25]提出了从时间维度基于多个流程的 SWRL 规则的方法, 用于威胁风险的判断。该方法的核心思想是将资产风险的分析过程分为多个步骤, 通过不同的过程定义 SWRL 规则, 最终将规则进行串联得到资产分析的结果。在具体实现中, 该方法将资产分析过程分为要素关联, 计算威胁似然, 识别受影响资产及程度, 分析威胁传播路径四个过程, 具有组织清晰, 逻辑严谨的特点。

文献[110]提出了从空间维度基于多个层次构建 SWRL 规则的方法, 用于识别错误的物联网安全配置。该方法的核心思想是通过不同层次的约束检查配置信息, 实现多个配置信息正确性的推理判断, 检测出存在风险的配置。在具体实现过程中将配置约束分为基础性约束(Fundamental Constraints)和用户驱动约束(User-driven Constraints)两个方面, 在不同的层次分别构建 SWRL 规则。其中, 基础性约束包括可达性约束、抽样约束、资源约束等针对内部信息的约束; 用户驱动约束包括能力约束和条件约束等外部信息的约束, 并且可以根据特定的攻击行为进一步扩展约束规则。

4.3.2　基于图的情报知识推理

基于图的推理借助图拓扑结构实现信息推理, 与基于规则的方法相比, 该推理过程更充分的利用了情报知识图谱中的关联信息, 具有表现力丰富、鲁棒性强、定义简单的特点。基于图的推理研究主要包括攻击图推理、社交网络推理和相似图判断推理。

攻击图是安全分析中常用的图形化结构, 具有丰富的研究[111]。攻击图使用有向图表示, 一般使用节点表示系统状态, 边表示可导致状态变化的攻击行为[112-113]。通过情报与系统状态的结合, 实现资产受威胁程度和潜在攻击路径的推理。攻击图方法一经提出便受到了广泛的关注, 但是在其发展过程中逐渐暴露了诸多问题, 主要可以归结为以下三方面: (1)如何有效提高大规模攻击图的推理效率; (2)如何定量分析潜在攻击路径的危害; (3)如何充分结合攻击图与知识。针对第一个问题, 存在两种解决方案: 一种方案是攻击图生成算法的优化。通过提高攻击图的逻辑性以及对攻击路径进行修剪, 达到减少计算复杂程度的目的, 提高攻击图的生成效率。例如, 文献[114]利用可达性分组的方法减少可达性矩阵的空间复杂度。文献[115]基于攻击概率对攻击图进行修剪以减少冗余的攻击路径。另一种方案是提高攻击图的计算能力。通过并行化的方法充分利用计算资源, 提高攻击图的计算能力, 从而减少攻击图的生成时间。例如, 文献[116]针对漏洞数量增加时攻击图状态爆炸式增长的问题提出了基于内存的并行分布算法, 可在分布式多代理平台上构建基于漏洞的攻击图, 从而减少攻击图的计算时间。针对第二个问题, 文献[117-119]提出了最短路径度量、路径数量度量和平均路径长度度量。但是, 最短路径忽略了其他攻击路径的危害, 路径数量度量未考虑单条路径的危害程度, 平均路径长度则无法表现攻击者可采用的路径数量。针对这个问题, 文献[120]提出了融合

路径数量和路径平均长度的正则化路径平均长度, 融合路径长度以及攻击者意图的标准路径长度偏差, 表示攻击路径长度频率的路径长度模式, 用于表示典型路径长度的路径长度中位数, 从而根据生成的攻击图定量评价资产的受攻击风险。此外, 文献[121]提出了基于本体的类别信息的聚类系数、度和连接类别差异的度量标准, 用于评价语义图关系发现的潜力。文献[122]引入马尔可夫链和漏洞生存周期模型, 分析漏洞随着存在时间的增长趋势以及相应补丁的发布对漏洞危害性的影响, 同时借助 CVSS(Common Vulnerability Scoring System, CVSS)漏洞评分系统, 实现对潜在攻击路径危害性的定量分析。针对第三个问题, 当前的研究思路是利用本体结构及本体上的规则生成攻击图, 实现知识与攻击图的结合。例如, 文献[123]将包含资产、漏洞、攻击的本体结构以及网络拓扑、受害主机等信息相结合构建了攻击图。在具体实现过程中, 通过使用推理算法 JESS[124]与 SWRL 规则不断迭代攻击过程, 从而识别受影响资产及其危害范围的分析。文献[125]提出了构建事件因果图的方法, 用于实现恶意软件的网络活动推理。该方法的核心思想是通过构建事件因果图, 检测无法回溯到正常系统行为的事件。在具体实现过程中, 首先定义了事件触发关系, 根据事件产生的原因, 将不同的事件关联起来构成触发图。然后在触发图中应用根触发策略判断网络活动是否存在正常的根节点。对于正常的网络事件, 可以追溯到产生事件的系统行为, 而对于恶意事件则无法追溯到合理的系统行为, 从而达到恶意行为识别的目的。

社交网络作为当前最受欢迎的网络沟通方式和信息传播媒介, 其上蕴含着海量的安全情报信息。因此, 近年来, 大量安全人员利用这些信息从事着安全事件分析和挖掘的研究。例如, 文献[126]提出了利用黑客论坛的信息推断手机恶意软件的关键传播者的方法。该方法的核心思想是通过发布恶意软件的行为对黑客进行组织关联, 同时识别恶意软件的关键传播人员。在具体实现过程中, 首先利用长短时效神经网络识别论坛附件是否为移动恶意软件, 然后通过构建单模式(one-mode)网络将恶意软件的发布者关联起来, 使用网络直径衡量黑客团体, 并根据共现性(co-occurrences)确定恶意软件的关键传播人员。该方法使用单模式网络更适合黑客团体性的挖掘, 同时可以简化网络中心性的计算。虽然该方法在跟进最新发布的恶意软件危害时具有良好的效果, 但是由于黑客论坛中缺少关于黑客的属性信息, 难以对黑客进行更为深入的挖掘。为了构建更为完整的用户画像, 文献[127-131]等提出了针对用户缺失属性的推理方法, 可以根据其社交信息进行推理, 主要可以分为基于好友和基于行为的两种攻击推理方法。文献[132]结合以上两种方法, 提出了 VIAL(Vote Distribution Attack)方法用于推理目标用户属性。该方法通过搜集 Facebook、Google Play 和 Google+等公开信息, 构建包含社交结构、属性信息、行为信息的 SBA 网络(Social Behavior Attribute, SBA network), 然后应用 VIAL 算法完成 SBA 网络中属性的权重分配过程, 实现针对目标用户的属性推理。

在一些研究中, 通过对不同图结构相似程度的判断达到信息推理的目的。相似图推理可用于恶意软件家族推理, 威胁情报检索等应用。相似图判断推理是在构建图拓扑结构之后, 通过图的相似度匹配进行关联推理等过程。在 APT 攻击中, 恶意软件通常以家族的方式进行演变, 因此建立恶意程序的家族图谱对于 APT 攻击的溯源以及对新的恶意程序的理解具有意义。例如, 文献[133]提出了基于结构先验和偏序先验知识的贝叶斯网络发现算法, 根据恶意样本的执行流图推理其家族的演进过程。但是该方法直接使用程序控制流图作为输入, 未对图结构进行简化, 在输入大量样本时将面临效率的问题。为了提高计算效率, 文献[134]提出了使用图嵌入算法将执行流图转化为向量表示并通过向量距离判断恶意代码相似成俗的方法。不同于研究较多的分类问题的图嵌入, 该方法的嵌入过程是为实现代码执行图的相似性检测, 而不是用于类别的判断。因此该方法在 Structure2Vec 的基础上, 使用 Siamese 结构实现无标签的参数化训练, 相对分类问题的嵌入而言保留了更多的原始特征, 对于无标签训练的恶意样本检测而言更为适用。同样基于简化图计算的思路, 文献[135]提出了将威胁信息的相关程度转化为威胁情报图相似程度判断的方法。该方法在统一数据模型(Unified Data Model)表示的基础上, 首先通过对齐不同层次的情报信息构成情报图, 而后使用相似哈希算法分别计算节点、子图与输入情报的指纹信息, 并通过海明距离衡量指纹之间的相似程度。

4.3.3 总结与讨论

知识推理在近几年随着知识图谱的发展取得了较大的进步[35, 51, 54, 136-137]。表 3 总结了情报知识推理代表性工作及方法特点的对应关系。在安全相关研究中, 基于规则的推理方法与基于本体的方法相结合, 充分利用了本体中蕴含的规则知识, 具有可靠和准确的特点。但是在高可靠性的同时带来了规

则使用范围局限的缺点, 为了覆盖较多的应用类型, 需要依靠分析人员制定大量的规则。基于图的推理方法利用图拓扑结构对于数据关系的丰富表现能力提高了信息的推理深度, 同时基于图的推理方法研究丰富, 但是计算效率较低是直接应用图方法面临的问题, 图拓扑结构的降维表示或许可以为这一问题的解决提供思路。除此之外, 近年在通用知识图谱推理中兴起的基于知识分布式表示的知识推理和基于深度学习的推理吸引了人们的关注[48]。基于知识分布式表示的推理首先使用嵌入式方法将知识的三元组表示转化为低维向量进行表示, 同时将推理操作变为基于向量的运算。知识的分布式表示方法主要可以分为基于转移距离模型的方法和基于语义匹配的方法[35]。基于分布式表示的方法具有运算简单, 推理效率高的特点, 但是难以涵盖复杂的语义逻辑, 推理能力有限。基于深度学习的推理期望利用神经网络刻画三元组间的语义联系, 并且有相应的研究开展[138-140]。通过模拟计算步骤或推理过程实现知识的推理, 具有表达能力丰富, 推理能力强的特点, 但是网络结构的可解释性仍是困扰神经网络研究进一步发展的难题。

5 讨论

目前, 对于情报知识图谱研究仍处在起步阶段, 在理论、模型以及具体构建过程等方面均存在着阻碍情报知识图谱发展与应用的问题, 同时对于新的发展方向需进一步深入研究, 本文从以下方面对情报知识图谱的研究进行探讨:

(1) 信息源的选择与维护

维护与情报研究相关的数据源, 是实现高效获取信息的前提。由于安全情报的领域专业性, 对于信息源既要求能广泛覆盖安全信息, 同时要减少安全无关信息的存在。Web 2.0 的发展使得信息呈现分散化趋势, 一次事件的细节往往无法通过一个信息源全部获得, 而是需要对多个资源进行综合才可得到事件的完整画像。但若是信息源中存在过多的安全无关信息, 则会耗费大量的时间处理无关数据, 亦会对情报的质量产生影响。

表 3 情报知识推理方法

Table 3 The Methods of Intelligence Knowledge Inference

类型	文献	推理内容	方法	优点	缺点
基于规则的情报知识推理	[108]	错误配置策略	基于 SQL 的逻辑判断	构建策略和行为的匹配模型, 通过 SQL 实现逻辑判断	SQL 的推理能力有限
	[109]	针对特定系统的告警	基于 SWRL 规则的推理	基于 UCO 本体进行规则构建	规则过于扁平化, 不利于规则的管理和扩充
	[25]	威胁风险	基于 SWRL 规则的推理	基于多个流程进行规则构建	不适用于简单流程的推理
	[110]	物联网安全配置	基于 SWRL 规则的推理	基于多个角度进行规则构建	规则的组织较为复杂, 耦合性与内聚性的平衡
基于图的情报知识推理	[126]	手机恶意软件传播者	社交网络分析	基于社交网络中心性的计算	未对恶意传播者深入挖掘
	[132]	用户属性	社交网络分析	属性推理攻击	需要大量的社交信息
	[123]	系统状态	攻击图推理	与 SWRL 规则相结合进行攻击图迭代	面对大规模网络时的效率问题
	[122]	系统风险	攻击图推理	引入漏洞生存周期	仅考虑漏洞对系统的影响
	[135]	威胁检索	相似图判断	将情报相似性转化为图的相似性	相似哈希算法的选取
	[125]	恶意软件网络活动	因果图推理	通过逻辑关系构建关联图	无法发现具有隐藏能力的攻击
	[133]	恶意软件行为	贝叶斯网络发现	融合结构先验和偏序先验知识	大量样本时的效率问题
	[134]	恶意软件行为	相似图判断	使用图嵌入的方法将图转化为向量	神经网络的训练需要大量的样本

(2) 信息抽取的质量

当前基于规则匹配的方法与基于统计学习的方法均难以满足面向开放域抽取的准确性、灵活性、鲁棒性的综合要求。构建知识图谱需要面临开放网络中的海量文本, 这些文本具有不同的结构, 同时涉及的安全情报信息在不断增加, 给信息抽取带来了较大困难。其中, 垂直语料的匮乏是信息抽取面临的较大问题。语料数据对于基于规则匹配的方法和基于统计学习的方法而言都是必不可少的元素。自然语言处理以及通用知识图谱的研究开始较早, 存

在大量的公开语料库, 相对来讲较为注重有监督方法的研究, 但是面向安全领域的垂直语料较为匮乏, 因此可着重考虑半监督抽取方法或无监督抽取方法的应用, 减少对标注垂直语料的依赖。

(3) 自动化本体构建

使用人工编辑的方法构建安全本体具有准确、易操作的优势, 但是自动化的本体构建仍有必要。在本体更新过程中, 需要及时收集新出现的知识, 同时关注知识体系的变化, 而这些变化会通过数据反映出来, 因此通过数据驱动的方式可以更高效的更新知识。自动化本体构建问题又称为本体学习(Ontology Learning), 当前, 已有的本体学习探索如下: (1)利用结构化信息直接作为先验知识。结构化数据中存在数据库模式、关系模式、结构层次等信息, 可以通过适当操作转化为本体[141-142]; (2)通过聚类方法从数据中生成新的概念及其上下文关系。层次聚类方法可以计算文本中实体的相似度, 将高相似度的实体集合组合成为新的概念, 同时提取概念的上下文关系, 进而通过本体语言描述转化为本体[143-146];

(4) 情报质量评估

情报质量评估是保证情报准确性和可用性的重要步骤。首先, 需要判断抽取的过程是否属于威胁情报的范畴。在面向开放域收集信息时, 会受到非安全情报信息的干扰, 因此需要对抽取信息的安全相关度进行判断。此外从多个信息源获取情报时, 由于信息源的来源的差异以及可靠性不同, 会对情报的融合造成困难, 突出表现在较多的冲突情报。针对这个问题, 可以考虑采用真值发现的方法确定正确的情报[147], 例如基于数据源质量的方法[148-150], 以及基于数据依赖关系的方法[151]。

(5)传统知识推理方法与安全应用的结合

通用推理方法具有丰富的研究, 但是这些方法主要面向三元组的推理, 与情报知识图谱的推理需求差别较大。如何将通用知识图谱的推理方法与业务应用结合, 达到解决现实安全问题的目的, 是安全推理研究面临的挑战。针对这一问题, 本文提出以下见解: 通过对业务逻辑的分解, 将通用知识图谱的推理方法应用到其中某些环节, 可以发挥出两者的优势。例如在基于图的推理中, 使用嵌入式方法将文字表示或图形表示为向量, 将基于图的运算转化为基于向量的计算, 从而提高计算的效率[152]。

(6)图片信息的转化保存

知识不仅可以用自然语言的方式进行保存, 有相当一部分知识使用图片的方式进行保存, 例如网络资产常使用拓扑图来表现网络架构及拓扑位置, APT 报告中使用图片来表现攻击流程以及各要素之间的关联关系。而如何实现针对图片信息的转化保存, 仍需进行探索研究。

6 结束语

近几年, 安全情报的研究和发展得到了学术界和工业界的广泛关注, 但其发展过程中面临着数据离散化分布、信息内容不实、情报综合分析困难等问题, 而知识图谱的出现为上述问题的解决提供了一个有效的解决方案。本文以通用知识图谱构建框架为基础, 结合安全领域的业务需求, 对知识图谱在安全情报领域的应用研究进行了分析定位, 介绍了可用于安全情报图谱构建的研究现状, 从信息抽取、本体构建、知识推理与应用三个关键过程进行了关键技术的梳理与总结, 并对存在的问题进行深入讨论并提出了分析建议。本文旨在将知识图谱技术引入安全情报研究中, 为情报领域的知识图谱构建提供框架及总结, 并给其他安全领域的研究提供一些借鉴。

致 谢 感谢中国科学院网络测评技术重点实验室的各位老师和同学提出的有益建议。感谢审稿专家和编辑部老师对本文提出的有益建议及指导。

参考文献

[1] M. K. Daly. Advanced persistent threat[J]. *Usenix, Nov,* 2009, 4(4): 2013-2016.

[2] J. R. Goodall, A. D'Amico, J. K. Kopylec. Camus: automatically mapping cyber assets to missions and users[C]. *Military Communications Conference, IEEE 2009. (MILCOM)*, 2009, 1-7.

[3] V. Kumar, J. Srivastava, A. Lazarevic. Managing cyber threats: issues, approaches, and challenges[M]. Springer Science & Business Media, 2006.

[4] E. W. Burger, M. D. Goodman, P. Kampanakis, et al. Taxonomy Model for Cyber Threat Intelligence Information Exchange Technologies[C]. *ACM Workshop on Information Sharing & Collaborative Security*, 2014, 51-60.

[5] P. Barford, M. Dacier, T. G. Dietterich, et al. Cyber SA: Situational Awareness for Cyber Defense[J]. *Cyber Situational Awareness,* 2010, 46: 3-13.

[6] S. Goel. Cyberwarfare: connecting the dots in cyber intelligence[J], *Communications of the Acm,* 2011, 54(8): 132-140.

[7] G. R. McMillan. Open Threat Intelligence. https://www.gartner.com/doc/2487216/definition-threat-intelligence. 2013.

[8] A. Mohaisen, O. Al-Ibrahim, C. Kamhoua, et al. Rethinking In-

formation Sharing for Threat Intelligence[J]. 2017,2(2):5-9.

[9] S. Barnum. Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX™)[J]. *Mitre Corporation,* 2014, 3(5):56-62.

[10] G. Farnham, K. Leune. Tools and standards for cyber threat intelligence projects[J]. *SANS Institute,* 2013, 3(2): 25-31.

[11] E. W. Burger, M. D. Goodman, P. Kampanakis, et al. Taxonomy model for cyber threat intelligence information exchange technologies[C]. *the 2014 ACM Workshop on Information Sharing & Collaborative Security(WISCS)*, 2014: 51-60.

[12] G. Settanni, Y. Shovgenya, F. Skopik, et al. Acquiring Cyber Threat Intelligence through Security Information Correlation[C]. *IEEE International Conference on Cybernetics(CYBCONF)*, 2017: 1-7.

[13] A. Caglayan, M. Toothaker, D. Drapeau, et al. Behavioral analysis of botnets for threat intelligence[J]. *Information systems and e-business management,* 2012, 10(4): 491-519.

[14] R. Lee. Threat Intelligence in an Active Cyber Defense (Part 2)[J]. *Retrieved from Recorded Future: https://www. recordedfuture. com/active-cyber-defense-part-1/Lee,* 2015, 5(6): 68-72.

[15] A. Singhal. Introducing the knowledge graph: things, not strings[J]. *Official google blog,* 2012, 6(9): 15-22.

[16] J. Lehmann, R. Isele, M. Jakob, et al. DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia[J]. *Semantic Web,* 2015, 6(2): 167-195.

[17] N. Shadbolt, T. Bernerslee, W. Hall. The Semantic Web Revisited[J]. *IEEE Intelligent Systems,* 2006, 21(3): 96-101.

[18] K. Bollacker, C. Evans, P. Paritosh, et al. Freebase:a collaboratively created graph database for structuring human knowledge[C]. in *SIGMOD Conference*, 2008, 1247-1250.

[19] S. Auer, C. Bizer, G. Kobilarov, et al. Dbpedia: A nucleus for a web of open data[M]. in *The semantic web*: Springer, 2007: 722-735.

[20] ZHIHU. 改变命运的知识，也会改变人工智能的发展轨迹. https://zhuanlan.zhihu.com/p/31846591. 2017.

[21] C. Jedrzejek, J. Bak, M. Falkowski. Graph mining for detection of a large class of financial crimes[C]. *International Conference on Conceptual Structures, Moscow, Russia*, 2009:124-131.

[22] 高. 漆桂林, 吴天星. 知识图谱研究进展[J]. *情报工程*, 2017, 3(1):004-025.

[23] Xu Zenglin , Sheng Yongpan, He lirong, et al. Review on Knowledge Graph Techniques[J], *Journal of University of Electronic Science and Technology of China*, 2016, (4): 589-606.
(徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报 2016, (4): 589-606.)

[24] Liu Jiao, Li Yang, Duan Hong, et al. Knowledge Graph Construction Techniques[J]. *Journal of Computer Research and Development*, 2016, 53(3): 582-600.
(刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. *计算机研究与发展*, 2016, 53(3): 582-600.)

[25] S. Qamar, Z. Anwar, M. A. Rahman, et al. Data-driven analytics for cyber-threat intelligence and information sharing[J]. *Computers & Security,* 2017, 67: 35-58.

[26] J. Friedman, M. Bouchard. Definitive guide to cyber threat intelligence[M]. *CyberEdge Press*, 2015.

[27] D. Chismon, M. Ruks. Threat intelligence: Collecting, analysing, evaluating[J]. *MWR InfoSecurity Ltd,* 2015, 3(2):36-42.

[28] A. Modi. CSM Automated Confidence Score Measurement of Threat Indicators[D]. Arizona State University, 2017.

[29] E. Nunes, A. Diab, A. Gunn, et al. Darknet and deepnet mining for proactive cybersecurity threat intelligence[C]. *Intelligence and Security Informatics(ISI)*, 2016, 7-12.

[30] Yang Zeming, Li Qiang, Liu Junrong, et al. Research of Threat Intelligence Sharing and Using for Cyber Attack Attribution[J]. *Journal of Information Security Research*, 2015, 1(01):37-42.
(杨泽明, 李强, 刘俊荣,等. 面向攻击溯源的威胁情报共享利用研究[J]. *信息安全研究*, 2015, 1(01):37-42.)

[31] WIKIPEDIA. Knowledge Graph. https://en.wikipedia.org/wiki/Knowledge_Graph. 2018.

[32] H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods[J]. *Semantic Web,* 2017, 8(3): 489-508.

[33] L. Zhang. Knowledge graph theory and structural parsing[J]. *University of Twente,* 2002, 2(4): 25-31.

[34] Y. Chen, Z. Lin, X. Zhao, et al. Deep Learning-Based Classification of Hyperspectral Data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing,* 2017, 7(6): 2094-2107.

[35] Q. Wang, Z. Mao, B. Wang, et al. Knowledge Graph Embedding: A Survey of Approaches and Applications[J]. *IEEE Transactions on Knowledge & Data Engineering,* 2017, 29(12): 2724-2743.

[36] G. Angeli, M. J. J. Premkumar, C. D. Manning. Leveraging linguistic structure for open domain information extraction[C]. *the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015: 344-354.

[37] P. S. Jacobs. Text-based intelligent systems: Current research and practice in information extraction and retrieval[M]. Psychology Press, 2014.

[38] E. Tsui, W. M. Wang, L. Cai, et al. Knowledge-based extraction of intellectual capital-related information from unstructured data[J]. *Expert Systems with Applications An International Journal,* 2014, 41(4): 1315-1325.

[39] H. Zhong, J. Zhang, Z. Wang, et al. Aligning Knowledge and Text Embeddings by Entity Descriptions[C]. *Conference on Empirical Methods in Natural Language Processing(EMNLP)*, 2015:

267-272.

[40] Wang Xuepeng, Liu Kang, He Shizhu, et al. Multi-Source Knowledge Base Entity Alignment by Leveraging Semantic Tags[J]. *Chinese Journal of Computers*, 2017, 040(003):701-711.
(王雪鹏, 刘康, 何世柱, 等. 基于网络语义标签的多源知识库实体对齐算法[J]. *计算机学报*, 2017, 040(003):701-711.)

[41] Zhuang Yan, Li Guoliang, Feng Jianhua. A Survey on Entity Alignment of Knowledge Base[J]. *Journal of Computer Research and Development*, 2016, 53(1): 165-192.
(庄严, 李国良, 冯建华. 知识库实体对齐技术综述[J]. *计算机研究与发展*, 2016, 53(1): 165-192.)

[42] T. R. Gruber. A translation approach to portable ontology specifications[J]. *Knowledge Acquisition,* 1993, 5(2): 199-220.

[43] J. N. K. Liu, Y. L. He, E. H. Y. Lim, et al. A New Method for Knowledge and Information Management Domain Ontology Graph Model[J]. *IEEE Transactions on Systems Man & Cybernetics Systems,* 2012, 43(1): 115-127.

[44] M. A. Storey, M. Musen, J. Silva, et al. Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protégé[J]. *in Protégé. Workshop on Interactive Tools for Knowledge Capture (K-CAP-2001,* 2001, 6(2):25-31.

[45] B. K. Foguem, T. Coudert, C. Béler, et al. Knowledge formalization in experience feedback processes: An ontology-based approach[J]. *Computers in Industry,* 2008, 59(7): 694-710.

[46] W. Wong, W. Liu, M. Bennamoun. Ontology learning from text:A look back and into the future[J]. *Acm Computing Surveys,* 2012, 44(4): 1-36.

[47] J. J. Miller. Graph database applications and concepts with Neo4j[C]. *the Southern Association for Information Systems Conference,* , 2013: 36.

[48] Guan Saiping, Jin Xiaolong, Jia Yantao, et al. Knowledge Graph Oriented Knowledge Inference Methods: A Survey[J]. *Journal of Software*, 2018, 29(10): 2966-2994.
(官赛萍, 靳小龙, 贾岩涛,等. 面向知识图谱的知识推理研究进展[J]. *软件学报*, 2018, 29(10): 2966-2994.)

[49] Klyne, Graham, Carroll, et al. Resource Description Framework (RDF): Concepts and Abstract Syntax[J]. *World Wide Web Consortium Recommendation,* 2004, 5(3): 261-265.

[50] Y. Luo, Q. Wang, B. Wang, et al. Context-Dependent Knowledge Graph Embedding[C]. *Conference on Empirical Methods in Natural Language Processing(EMNLP)*, 2015: 1656-1661.

[51] Z. Wang, J. Zhang, J. Feng, et al. Knowledge Graph Embedding by Translating on Hyperplanes[J]. *AAAI - Association for the Advancement of Artificial Intelligence,* 2014, 5(2): 125-134.

[52] G. Ji, S. He, L. Xu, et al. Knowledge Graph Embedding via Dynamic Mapping Matrix[C]. *Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, 2015: 687-696.

[53] Y. Lin, Z. Liu, X. Zhu, et al. Learning entity and relation embeddings for knowledge graph completion[C]. *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015: 2181-2187.

[54] A. Bordes, N. Usunier, A. Garcia-Duran, et al. Translating Embeddings for Modeling Multi-relational Data[C]. *International Conference on Neural Information Processing Systems(NIPS)*, 2013: 2787-2795.

[55] J. Cowie, W. Lehnert. Information extraction[J]. *Communications of the ACM,* 1996, 39(1): 80-91.

[56] S. Soderland. Learning Information Extraction Rules for Semi-Structured and Free Text[J]. *Machine Learning,* 1999, 34(1-3): 233-272.

[57] M. E. Califf, R. J. Mooney. Relational learning of pattern-match rules for information extraction[C]. *Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence*, 1999: 328-334.

[58] L. Chiticariu, Y. Li, F. R. Reiss. Rule-based information extraction is dead! long live rule-based information extraction systems![C]. in *Proceedings of the 2013 conference on empirical methods in natural language processing(EMNLP)*, 2013: 827-832.

[59] Zhao Jun, Liu Kang, Zhou Guangyou, et al. Open Information Extraction[J]. *Journal of Chinese Information Processing*, 2011, 25(6):98-111.
(赵军, 刘康, 周光有,等. 开放式文本信息抽取[J]. *中文信息学报*, 2011, 25(6):98-111.)

[60] S. Z. Zhang, H. Luo, B. X. Fang. Regular Expressions Matching for Network Security[J]. *Journal of Software,* 2011, 22(8): 1838-1854.

[61] S. Kushner. Ontology-Driven Data Semantics Discovery for Cyber-Security[C]. *Practical Aspects of Declarative Languages(PADL)*, 2015: 1-16.

[62] X. Liao, K. Yuan, Z. Li, et al. Acing the IOC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence[C]. *ACM Sigsac Conference on Computer and Communications Security(ACM SIGSAC)*, 2016: 755-766.

[63] M. Thelen, E. Riloff. A bootstrapping method for learning semantic lexicons using extraction pattern contexts[C]. *Acl-02 Conference on Empirical Methods in Natural Language Processing(EMNLP)*, 2002: 214-221.

[64] H. Yu. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences[J]. *Proceedings of Emnlp,* 2003, 116(3): 129-136.

[65] J. Betteridge, A. Carlson, S. A. Hong, et al. Toward Never Ending Language Learning[C]. *.AAAI spring symposium: Learning by reading and learning to read*, 2009: 1-2.

[66] C. L. Jones, R. A. Bridges, K. M. T. Huffer, et al. Towards a Relation Extraction Framework for Cyber-Security Concepts[C]. *Cyber and Information Security Research Conference(CISR)*, 2015: 11.

[67] N. Bach, S. Badaskar. A SURVEY ON RELATION EXTRACTION[J], *Language Technologies Institute,* 2007, 23(3): 268-271.

[68] N. McNeil, R. A. Bridges, M. D. Iannacone, et al. Pace: Pattern accurate computationally efficient bootstrapping for timely discovery of cyber-security concepts[C]. *Machine Learning and Applications(ICMLA)*, 2013: 60-65.

[69] A. Ratnaparkhi. Maximum entropy models for natural language ambiguity resolution[J], 1998, 25(2): 369-371.

[70] S. Zheng, S. Jayasumana, B. Romera-Paredes, et al. Conditional random fields as recurrent neural networks[C]. *the IEEE International Conference on Computer Vision(ICCV)*, 2015: 1529-1537.

[71] J. Yoo, H. H. Kwon, B. J. So, et al. Identifying the role of typhoons as drought busters in South Korea based on hidden Markov chain models[J]. *Geophysical Research Letters,* 2015, 42(8): 2797-2804.

[72] B. Ramesh, C. Xiang, T. H. Lee. Shape classification using invariant features and contextual information in the bag-of-words model[J]. *Pattern Recognition,* 2015, 48(3): 894-906.

[73] V. Mulwad, W. Li, A. Joshi, et al. Extracting Information about Security Vulnerabilities from Web Text[C]. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2011: 257-260.

[74] S. More, M. Matthews, A. Joshi, et al. A Knowledge-Based Approach to Intrusion Detection Modeling[C], *IEEE Symposium on Security and Privacy Workshops(SP)*, 2012, 75-81.

[75] J. R. Finkel, T. Grenager, C. Manning. Incorporating non-local information into information extraction systems by Gibbs sampling[C], *Meeting on Association for Computational Linguistics(ACL)*, 2005: 363-370.

[76] B. Settles. ABNER: an open source tool for automatically tagging genes, proteins and other entity names in text[J]. *Bioinformatics,* 2005, 21(14): 3191-3192.

[77] R. Lal. Information Extraction of Security related entities and concepts from unstructured text[J]. 2013, 26(5): 263-265.

[78] A. Joshi, R. Lal, T. Finin, et al. Extracting Cybersecurity Related Linked Data from Text[C]. *IEEE Seventh International Conference on Semantic Computing(ICSC)*, 2013: 252-259.

[79] R. Lal. Information Extraction of Security related entities and concepts from unstructured text[J]. *Dissertations & Theses - Gradworks,* 2013, 2(3): 698-670.

[80] R. A. Bridges, C. L. Jones, M. D. Iannacone, et al. Automatic Labeling for Entity Extraction in Cyber Security[J], *Computer Science,* 2013: 258-261.

[81] R. Studer, V. R. Benjamins, D. Fensel. Knowledge engineering: Principles and methods[J]. *Data & Knowledge Engineering,* 1998, 25(1–2): 161-197.

[82] L. Drumond, R. Girardi. A Survey of Ontology Learning Procedures[J]. *WONTO,* 2008, 427: 1-13.

[83] V. Mavroeidis, S. Bromander. Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence[C]. *Intelligence and Security Informatics Conference(ISI)*, 2017: 91-98.

[84] M. M. Kokara, C. J. Matheusb, K. Baclawskic. Ontology-based situation awareness[J]. *Information Fusion,* 2009, 10(1): 83-98.

[85] J. A. Wang, M. Guo. OVM: an ontology for vulnerability management[C], *the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies*, 2009: 34.

[86] A. Patel, M. Taghavi, K. Bakhtiyari, et al. An intrusion detection and prevention system in cloud computing: A systematic review[J]. *Journal of Network & Computer Applications,* 2013, 36(1): 25-41.

[87] L. D. Xu, W. He, S. Li. Internet of Things in Industries: A Survey[J]. *IEEE Transactions on Industrial Informatics,* 2014, 10(4): 2233-2243.

[88] J. Weston, A. Bordes, O. Yakhnenko, et al. Connecting language and knowledge bases with embedding models for relation extraction[EB/OL]. 2013: *ArXiv Preprint ArXiv:1307.7973,* 2013.

[89] A. Oltramari, L. F. Cranor, R. J. Walls, et al. Building an Ontology of Cyber Security[C], *Semantic Technology for Intelligence Defense and Security(STIDS)*, 2014: 54-61.

[90] A. Oltramari, D. Henshel, M. Cains, et al. Towards a Human Factors Ontology for Cyber Security[C]. *Semantic Technology for Intelligence Defense and Security(STIDS)*, 2015: 258-267.

[91] C. L. Maines, D. Llewellyn-Jones, S. Tang, et al. A Cyber Security Ontology for BPMN-Security Extensions[C]. *IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015: 1756-1763.

[92] S. Fenz, A. Ekelhart. Formalizing information security knowledge[C]. *the 4th international Symposium on information, Computer, and Communications Security*, 2009: 183-194.

[93] B. Guttman, E. A. Roback. An introduction to computer security: the NIST handbook[M]. DIANE Publishing, 1995.

[94] I. BSI. Grundschutz manual[M]. BSI, 2004.

[95] M. Pendleton, R. Garcia-Lebron, J.-H. Cho, et al. A survey on systems security metrics[J]. *ACM Computing Surveys (CSUR),* 2017, 49(4): 62.

[96] M. Atighetchi, B. I. Simidchieva, F. Yaman, et al. Using Ontologies to Quantify Attack Surfaces[C]. *Semantic Technology for Intelligence Defense and Security(STIDS)*, 2016: 10-18.

[97] N. Ben-Asher, S. Hutchinson, A. Oltramari. Characterizing network behavior features using a cyber-security ontology[C]. *Military Communications Conference 2016(MILCOM)*, 2016: 758-763.

[98] A. Vorobiev, N. Bekmamedova. An Ontology-Driven Approach Applied to Information Security[J]. *Journal of Research & Practice in Information Technology,* 2010, 42(1): 61-76.

[99] F. Mouton, L. Leenen, H. S. Venter. Social engineering attack examples, templates and scenarios[J]. *Computers & Security,* 2016, 59: 186-209.

[100] M. Iannacone, S. Bohn, G. Nakamura, et al. Developing an Ontology for Cyber Security Knowledge Graphs[C]. *Cyber and Information Security Research Conference(CISRC)*, 2015: 12.

[101] Z. Syed, A. Padia, T. Finin, et al. UCO: A Unified Cybersecurity Ontology[C], *AAAI Workshop on Artificial Intelligence for Cyber Security*, 2016: 259-261.

[102] L. Obrst, P. Chase, R. Markeloff. Developing an Ontology of the Cyber Security Domain[C]. *Semantic Technology for Intelligence Defense and Security(STIDS)*, 2012: 49-56.

[103] WIKIPEDIA. Knowledge Inference. https://zh.wikipedia.org/wiki/%E6%8E%A8%E7%90%86. 2018.

[104] T. C. Chiang, C. F. Tai, T. W. Hou. A knowledge-based inference multicast protocol using adaptive fuzzy Petri nets[J]. *Expert Systems with Applications,* 2009, 36(4): 8115-8123.

[105] S. Caltagirone, A. Pendergast, C. Betz. The diamond model of intrusion analysis[DB]. CENTER FOR CYBER INTELLIGENCE ANALYSIS AND THREAT RESEARCH HANOVER MD, 2013.

[106] E. M. Hutchins, M. J. Cloppert, R. M. Amin. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains[J]. *Leading Issues in Information Warfare & Security Research,* 2011, 1(1): 80.

[107] S. Bechhofer. OWL: Web ontology language[M]. *Encyclopedia of database systems*: Springer, 2009: 2008-2009.

[108] A. M. Azab, A. M. Azab, W. Enck, et al. SPOKE: Scalable Knowledge Collection and Attack Surface Analysis of Access Control Policy for Security Enhanced Android[C]. *ACM on Asia Conference on Computer and Communications Security(ASIACCS)*, 2017, 612-624.

[109] S. Mittal, P. K. Das, V. Mulwad, et al. CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities[C]. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining(ASONAM)*, 2016, 860-867.

[110] M. Mohsin, Z. Anwar, F. Zaman, et al. IoTChecker : a data-driven framework for security analytics of internet of things configurations[J], *Computers & Security,* 2017, 24(3):70.

[111] K. Durkota, V. Lisý, B. Bosanský, et al. Optimal Network Security Hardening Using Attack Graph Games[C]. *International Joint Conferences on Artificial Intelligence(IJCAI)*, 2015: 526-532.

[112] Liu Weixin, Zheng Kangfeng, Wu Bin, et al. Alert Processing Based on Attack Graph and Multi-source Analyzing[J]. *Journal of Communications*, 2015(09): 135-144.
(刘威歆, 郑康锋, 武斌, 等. 基于攻击图的多源告警关联分析方法[J]. *通信学报*, 2015(09): 135-144.)

[113] Ye Ziwei, Guo Yuanbo, Wang Chendong, et al. Survey on Application of Attack Graph Technology[J], *Journal of Communications*, 2017, 38(011): 121-132.
(叶子维, 郭渊博, 王宸东,等. 攻击图技术应用研究综述[J]. *通信学报*, 2017, 38(011): 121-132.)

[114] K. Ingols, R. Lippmann, K. Piwowarski. Practical Attack Graph Generation for Network Defense[C]. *Annual Computer Security Applications Conference(ACSAC)*, 2006: 121-130.

[115] A. Xie, L. Zhang, J. Hu, et al. A probability-based approach to attack graphs generation[C]. *International Symposium on Electronic Commerce and Security(ISESC)*, 2009: 343-347.

[116] K. Kaynar, F. Sivrikaya, Distributed Attack Graph Generation[J], *IEEE Transactions on Dependable & Secure Computing,* 2016, 13(5): 519-532.

[117] C. Phillips, L. P. Swiler. A graph-based system for network-vulnerability analysis[C]. *the 1998 workshop on New security paradigms*, 1998: 71-79.

[118] R. Ortalo, Y. Deswarte, M. Kaaniche. Experimenting with quantitative evaluation tools for monitoring operational security[J]. *IEEE Transactions on Software Engineering,* 2002, 25(5): 633-650.

[119] W. Li, R. B. Vaughn. Cluster Security Research Involving the Modeling of Network Exploitations Using Exploitation Graphs[C]. *IEEE International Symposium on Cluster Computing and the Grid(CCGRID)*, 2006: 26.

[120] N. Idika, B. Bhargava. Extending Attack Graph-Based Security Metrics and Aggregating Their Application[J]. *IEEE Transactions on Dependable & Secure Computing,* 2011, 9(1): 75-85.

[121] M. Barthelemy, E. Chow, T. Eliassi-Rad. Knowledge Representation Issues in Semantic Graphs for Relationship Detection[EB/OL]. 2005: *Arxiv Preprint Arxiv:* cs/0504072.

[122] S. Abraham, S. Nair. A Predictive Framework for Cyber Security Analytics using Attack Graphs[J]. *International Journal of Computer Networks & Communications,* 2015, 7(1): 269-271.

[123] S. Wu, Y. Zhang, W. Cao. Network security assessment using a semantic reasoning and graph based approach [J]. *Computers & Electrical Engineering,* 2017, 26(2): 64.

[124] E. F. Hill. Jess in action: Java rule-based systems[M]: Manning Publications Co., 2003.

[125] H. Zhang, D. Yao, N. Ramakrishnan, et al. Causality reasoning about network events for detecting stealthy malware activities 1[J]. *Computers & Security,* 2016, 58(C): 180-198.

[126] J. Grisham, S. Samtani, M. Patton, et al. Identifying mobile mal-

ware and key threat actors in online hacker forums for proactive cyber threat intelligence[C]. *IEEE International Conference on Intelligence and Security Informatics(ISI)*, 2017: 13-18.

[127] N. Z. Gong, A. Talwalkar, L. Mackey, et al. Joint Link Prediction and Attribute Inference Using a Social-Attribute Network[J]. *Acm Transactions on Intelligent Systems & Technology,* 2014, 5(2): 27.

[128] J. He, W. W. Chu, Z. Liu. Inferring Privacy Information from Social Networks[C]. *IEEE International Conference on Intelligence and Security Informatics(ISI)*, 2006: 154-165.

[129] J. Lindamood, R. Heatherly, M. Kantarcioglu, et al. Inferring private information using social network data[C]. *International Conference on World Wide Web(WWW)*, 2013: 1145-1146.

[130] K. Thomas, C. Grier, D. M. Nicol. unfriendly: multi-party privacy risks in social networks[C]. *International Conference on Privacy Enhancing Technologies Symposium(PETS)*, 2010: 236-252.

[131] E. Zheleva, L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles[C]. *International Conference on World Wide Web(WWW)*, 2009: 531-540.

[132] N. Z. Gong, B. Liu. You Are Who You Know and How You Behave: Attribute Inference Attacks via Users' Social Friends and Behaviors[C]. *USENIX Security Symposium(USENIX)*, 2016: 979-995.

[133] D. Oyen, B. Anderson. C. M. Anderson-Cook, Bayesian Networks with Prior Knowledge for Malware Phylogenetics[C]. *AAAI Workshop: Artificial Intelligence for Cyber Security*, 2016:58-64.

[134] X. Xu, C. Liu, Q. Feng, et al. Neural Network-based Graph Embedding for Cross-Platform Binary Code Similarity Detection[C]. *the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017: 363-376.

[135] H. Gascon, B. Grobauer, T. Schreck, et al. Mining Attributed Graphs for Threat Intelligence[C]. *ACM on Conference on Data and Application Security and Privacy(CODSPY)*, 2017: 15-22.

[136] G. Shu, W. Quan, L. Wang, et al. Jointly Embedding Knowledge Graphs and Logical Rules[C]. *Conference on Empirical Methods in Natural Language Processing(EMNLP)*, 2016: 192-202.

[137] H. Liu, Y. Wu, Y. Yang. Analogical inference for multi-relational embeddings[EB/OL]. 2017: *ArXiv Preprint ArXiv:1705.02426.*

[138] R. Das, A. Neelakantan, D. Belanger, et al. Chains of reasoning over entities, relations, and text using recurrent neural networks[EB/OL]. 2016: *ArXiv Preprint ArXiv:1607.01426.*

[139] A. Graves, G. Wayne, M. Reynolds, et al. Hybrid computing using a neural network with dynamic external memory[J]. *Nature,* 2016, 538(7626): 471-476.

[140] R. Socher, D. Chen, C. D. Manning, et al. Reasoning with neural tensor networks for knowledge base completion[C]. *International Conference on Neural Information Processing Systems(NIPS)*, 2013: 926-934.

[141] V. Kashyap. Design and creation of ontologies for environmental information retrieval[C]. *the 12th Workshop on Knowledge Acquisition, Modeling and Management*, 1999: 1-18.

[142] J. Lehmann, P. Hitzler. A Refinement Operator Based Learning Algorithm for the $\mathcal {ALC} $ Description Logic[C]. in *International Conference on Inductive Logic Programming(ILC)*, 2007: 147-160.

[143] G. Bisson, C. Nédellec, D. Canamero. Designing Clustering Methods for Ontology Building-The Mo'K Workbench[C]. *ECAI workshop on ontology learning*, 2000: 589-561.

[144] C. Wang, M. Danilevsky, N. Desai, et al. A phrase mining framework for recursive construction of a topical hierarchy[C]. *the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013: 437-445.

[145] X. Liu, Y. Song, S. Liu, et al. Automatic taxonomy construction from keywords[C]. *the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012: 1433-1441.

[146] E. Drymonas, K. Zervanou, E. G. Petrakis. Unsupervised ontology acquisition from plain texts: the OntoGain system[C]. *International Conference on Application of Natural Language to Information Systems*, 2010: 277-287.

[147] Y. Li, J. Gao, C. Meng, et al. A survey on truth discovery[J]. *ACM Sigkdd Explorations Newsletter,* 2016, 17(2): 1-16.

[148] X. Li, X. L. Dong, K. B. Lyons, et al. Scaling up copy detection[C]. *IEEE International Conference on Data Engineering(ICDE)*, 2015: 89-100.

[149] Ma Ruxia, Meng Xiaofeng, Wang Lu, et al. MTruths: An Approach of Multiple Truths Finding from Web Information[J]. *Journal of Computer Research and Development*, 2016, 52(012): 2858-2866. (马如霞, 孟小峰, 王璐, 等. MTruths:Web 信息多真值发现方法[J]. *计算机研究与发展*, 2016, 52(012): 2858-2866.)

[150] F. Zhang, L. Yu, X. Cai, et al. Truth finding from multiple data sources by source confidence estimation[C]. *Web Information System and Application Conference (WISA)*, 2015: 153-156.

[151] M. Lamine Ba, R. Horincar, P. Senellart, et al. Truth finding with attribute partitioning[C]. *the 18th International Workshop on Web and Databases(WebDB)*, 2015: 27-33.

[152] A. Roy, Y. Park, S. Pan. Learning Domain-Specific Word Embeddings from Sparse Cybersecurity Texts[EB/OL]. 2017: *ArXiv Preprint ArXiv:1709.07470.*

董聪 于2017年在天津大学信息管理与信息系统(保密方向)专业获得学士学位。现在中国科学院信息工程研究所第六研究室攻读博士学位。研究领域为网络安全态势感知、知识图谱等。Email: dongcong@iie.ac.cn

姜波 于 2016 年在中国科学院大学计算机系统结构专业获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为网络安全态势感知、知识图谱、数据挖掘等。Email: jiangbo@iie.ac.cn

卢志刚 于 2010 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所高级工程师, 中国科学院网络空间安全学院副教授。研究领域为网络安全态势感知、网络攻击检测、移动终端安全等。Email: luzhigang@iie.ac.cn

刘宝旭 于 2002 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所研究员, 第六研究室主任。研究领域为网络安全攻防对抗、网络安全测评技术等。Email: liubaoxu@iie.ac.cn

李宁 于 2014 年在中国科学院计算技术研究所获得博士学位。现任中国科学院信息工程研究所助理研究员。研究领域为网络安全态势感知。Email: lining6@iie.ac.cn

马平川 于 2017 年在大连理工大学计算机科学与技术专业获得学士学位。现在中国科学院信息工程研究所第六研究室攻读硕士学位。研究领域为网络安全态势感知、自然语言处理等。Email: mapingchuan@iie.ac.cn

姜政伟 于 2014 年在中国科学院大学获得博士学位。现任中国科学院信息工程研究所高级工程师, 中国科学院网络空间安全学院副教授。研究领域为威胁情报、态势感知、网络威胁发现。Email: jiangzhengwei@iie.ac.cn

刘俊荣 于2010年在北京邮电大学获得硕士学位, 现任中国科学院信息工程研究所高级工程师。研究领域为网络安全态势感知, 网络安全可视化等。Email: liujunrong@iie.ac.cn