基于网络行为的攻击同源分析方法研究

白 波[1,2,3], 冯 云[1], 刘宝旭[1,3], 汪旭童[1,3], 何松林[1,3], 姚敦宇[1,3], 刘奇旭[1,3]

[1]中国科学院信息工程研究所 北京 中国 100093
[2]北京网络数据研究所 北京 中国 100084
[3]中国科学院大学网络空间安全学院 北京 中国 100049

摘要 网络攻击威胁日益严峻, 攻击溯源是增强防御能力、扭转攻防局势的重要工作,攻击的同源分析是溯源的重要环节, 成为研究热点。根据线索类型的不同, 攻击同源分析可以分为基于恶意样本的同源分析和基于网络行为的同源分析。目前基于恶意样本的同源分析已经取得了较为显著的研究成果, 但存在一定的局限性, 不能覆盖所有的攻击溯源需求, 且由于恶意代码的广泛复用情况, 使得分析结果不一定可靠; 相比之下, 基于网络行为的同源分析还鲜有出色的成果, 成为溯源工作的薄弱之处。为解决现存问题, 本文提出了一种基于网络行为的攻击同源分析方法, 旨在通过抽取并分析攻击者或攻击组织独特的行为模式而实现更准确的攻击同源。为保留攻击在不同阶段的不同行为特征, 将每条攻击活动划分为 5 个攻击阶段, 然后对来自各 IP 的攻击行为进行了 4 个类别共 14 个特征的提取, 形成行为特征矩阵, 计算两两 IP 特征矩阵之间的相似性并将其作为权值构建 IP 行为网络图, 借助社区发现算法进行攻击社区的划分, 进而实现攻击组织的同源分析。方法在包含 114,845 条告警的真实的数据集上进行了实验, 凭借实际的攻击组织标签进行结果评估, 达到 96%的准确率, 证明了方法在攻击同源分析方面的有效性。最后提出了未来可能的研究方向。

关键词 攻击同源; 网络行为; 社区发现; 高级持续性威胁
中图法分类号 TP393.0 DOI 号 10.19363/J.cnki.cn10-1380/tn.2023.03.06

Research on Network Behavior-based Cyberattack Grouping Method

BAI Bo[1,2,3], FENG Yun[1], LIU Baoxu[1,3], WANG Xutong[1,3], HE Songlin[1,3], YAO Dunyu[1,3], LIU Qixu[1,3]

[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[2]Beijing Institute of Network Data, Beijing 100084, China
[3]School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The threat of cyberattacks is becoming more and more serious. Cyberattack attribution is a significant work to enhance defense capability and reverse the situation of attack and defense. Attack grouping analysis is an important part of attack attribution and has become a research hotspot. According to different types of clues, attack grouping analysis can be divided into grouping analysis based on malware and grouping analysis based on network behavior. At present, grouping analysis based on malware has achieved remarkable research results, but there are some limitations, which cannot cover all the requirements of attack attribution, and the analysis results are not necessarily reliable due to the widely reuse of malicious code. In contrast, grouping analysis based on network behavior has few outstanding results, which has become the weakness of attack attribution. In order to solve the existing problems, this paper proposes an attack grouping analysis method based on network behavior, which aims to achieve more accurate attack grouping by extracting and analyzing the unique behavior patterns of attackers or attack organizations. In order to retain the different behavioral characteristics of the attack in different stages, one attack activity is recognized into five attack stages, and then a total of 14 features of four categories are extracted from the attack behavior of each IP to form the behavior feature matrix. Then, calculate the similarity between every two IP feature matrices, and treat them as weights to construct the IP behavior network diagram. By using the community discovery algorithm, the attack community is divided, and then the grouping analysis of attack organizations is realized. The experiments were conducted on real datasets which include 114,845 warnings. The results were evaluated with the actual attack organization tags, and the accuracy was 96%, which proved the effectiveness of the method in attack homology analysis. Finally, the possible research directions in the future are put forward.

Key words attack grouping; network behavior; community discovery; APT

通讯作者: 刘宝旭, 工学博士, 研究员, Email: liubaoxu@iie.ac.cn。
本课题得到中国科学院青年创新促进会(No.2019163), 国家自然科学基金项目(No. 61902396), 中国科学院战略性先导科技专项项目(No. XDC02040100)课题资助; 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资助。
收稿日期: 2021-11-30; 修改日期: 2022-02-19; 定稿日期: 2023-01-04

1 引言

网络空间安全关系着国家民生活动、经济活动、军事活动、科学研究和政府决策的安全, 关系到各类企业、科研院所、政府机关等实体机构的切身利益。根据国家互联网应急中心(CNCERT)的《2020 年中国互联网网络安全报告》[1]统计, 我国全年捕获恶意样本数量超过 4200 万个, 境内受恶意程序攻击的 IP 地址约 5541 万个, 攻击手段持续升级, 对抗不断加剧。这些不断增加的网络攻击也呈现出高针对性、高复杂性、高组织性的演变趋势, 即高级持续性威胁(Advanced Persistent Threat, APT)。随着 APT 攻击的频次和范围明显增加, 这些组织的高级攻击技巧和思路也对普通的网络攻击者起到了很大的推波助澜作用, 部分刚出现的 APT 攻击技术和工具在一段时间后会被普通的网络攻击行动采用, 针对重点单位的 APT 攻击最终很可能会殃及普通网民。如有一起典型的 APT 攻击案例, 某黑客团队对驱动人生的公司发起了定向攻击, 采用了供应链攻击方式, 通过控制、篡改服务器配置, 利用正常软件的升级通道大规模安装远控木马[2]; 又如著名的 APT 武器“永恒之蓝”, 在被 ShadowBroker 泄露之后被众多普通攻击者制作成包括 WannaCry 在内的大量恶意软件, 至今仍在网络空间肆虐。作为安全研究人员, 既需要防范伪装性、隐蔽性都很强的 APT 攻击, 也需要应对愈演愈烈的其他网络攻击, 对攻击进行同源分析, 做到“知己知彼”。

在这样的情况下, 已知的攻击组织持续活跃, 未知的攻击组织不断涌现, 网络攻击态势更加严峻[3]。为了应对这样的态势, 对网络攻击的发现和溯源逐渐成为网络空间安全研究工作的重点需求。大多成规模的实体机构均设有自身的网络安全分析团队以提升在网络空间的防御能力, 这些团队的主要职责就是发现、防御和处置来自网络空间的各类攻击活动, 包括普通网络攻击和 APT 攻击。为了实现可持续的防御能力, 安全分析人员除了及时掌握更准确、更新的攻击手法并转换成检测技术之外, 还需要对检测结果中大量的攻击活动进行归类和同源分析, 支撑对攻击的溯源工作。攻击溯源工作的意义有以下几点:

(1) 能够发现隶属于已知攻击组织的新活动, 协助防御团队更新已掌握的攻击组织技战术情报。

(2) 能够从大量攻击告警中分离出有组织的行为和无组织的行为, 并进一步分析出已知攻击组织和潜在攻击组织, 支撑后续的溯源工作。

(3) 提高网络防御工作的针对性, 提高安全团队对于不同来源的有组织的攻击行动的技战术掌握程度, 配合用户行为分析等技术手段能够做到“知己知彼”, 增强机构实体的综合防御能力。

传统的攻击追踪溯源技术如路由调试[4]、数据包标记[5-6]等方法主要针对具有明显流量特征的网络攻击, 对先验知识要求较高。在网络空间的现实情况中, 充斥着大量商用和开源攻击工具造成的行为, 这些行为通常由于代码和功能的雷同而具有相似的流量特征, 因此上述基于明显流量特征的溯源和同源分析技术的现实应用效果会受到较大影响。目前有关网络攻击行为的学术研究更多的集中于攻击检测和发现[7-9], 同源分析工作较少, 由于网络数据的敏感性和隐私性, 攻击行为数据往往难以获取, 同源分析和溯源结果难以证明, 在一定程度上造成了目前相关学术研究的匮乏。为了解决这个困难, 本文参考 2017 年 UC Berkely 和 The Lawrence Berkeley National Laboratory 在 USENIX 发表的工作[10], 采用了从合作方的真实网络数据中提取流量数据作为数据集进行实验, 并根据国家网络空间威胁情报共享开放平台(CNTIC)和国家互联网应急中心(CNCERT)的数据进行富化, 使本文的方法验证和应用都基于真实世界中的网络数据, 更具有现实意义。

攻击行为的同源分析是进行网络攻击溯源的必要环节, 由于网络攻击的隐蔽性和匿名性, 鲜有攻击者能够直接通过行为特征等线索暴露身份, 因此, 攻击同源性分析, 即分析不同的网络攻击是否具有组织性、是否源自同一攻击者或攻击组织, 成为了定位攻击源、揭露攻击者身份的重要研究方向。对于攻击源头的直接追溯涉及到较为复杂的法律授权问题, 而同源分析相较于直接追溯攻击源头, 在数据和结论证明上更具可行性, 也使得学术研究可以有的放矢。

目前大多安全企业在安全服务工作的基础上均设有专门的威胁研究部门, 持续开展威胁追踪溯源工作, 并不定期发布各类溯源报告[11-13]。这些开展威胁追踪溯源的团队多是利用威胁情报技术辅助同源分析[14-16], 通过广泛收集已知的攻击组织信息构建情报库, 对从攻击中获取到的攻击样本或行为日志进行关联扩线, 匹配到相关的情报从而判断攻击源, 但也受限于情报信息的准确性和时效性, 更容易受到攻击者刻意伪装等因素的影响。例如当前在网络黑产的攻击者中存在一种伪装方式, 即普通攻击团伙通过续租知名 APT 组织已被威胁情报标注的过期域名, 将自己伪装成知名 APT 组织, 用以隐藏自己

的身份和攻击目的, 迷惑受攻击目标的安全团队。

根据线索类型的不同, 已有的同源分析工作可以分为基于样本的同源分析和基于行为的同源分析。基于样本的同源分析工作围绕攻击者使用的恶意代码展开研究, 通过逆向分析、动态监测等手段发现攻击设施、提取攻击者的编码风格和个性化特征。基于行为的同源分析围绕攻击者的攻击行为和网络日志展开研究, 从中抽取攻击行为模式。样本中的静态线索很容易被攻击者察觉从而刻意清除或伪造, 恶意代码在黑色产业链中也广泛存在复用现象, 因此对于大多数非特异性的恶意代码, 样本同源分析无法保证对真正攻击源同源分析结果的可靠性。但攻击行为受到攻击者/攻击组织的目的意图、攻击习惯、财务状况、人员组成、地域分布和其他潜在因素的影响, 因此存在固有的隐蔽行为模式, 难以被刻意伪装。本文提出一种基于网络行为的攻击同源分析技术, 通过分析告警及其对应的网络行为日志, 提取多维行为特征, 刻画攻击模式, 进而分析哪些行为属于同一攻击者或攻击组织, 实现基于网络行为的攻击同源分析。

本文的具体贡献如下:

(1) 通过对网络攻击行为的深度分析, 提取了能够表征攻击行为模式的攻击同源相关特征, 包含攻击目标、攻击设施属性、活动规律、个性化特点 4 个类别的共 14 个特征;

(2) 设计了攻击同源分析模型, 利用社区发现算法将不同的攻击行为划分到不同的攻击源社区, 实现攻击行为同源;

(3) 基于包含 114,845 条告警的真实攻击数据集进行了实验验证和评估, 并对实验结果做案例分析, 讨论了模型和方法的有效性。

其余章节安排: 第 2 节介绍相关研究现状; 第 3 节阐述本文所提出的网络行为同源分析方法; 第 4 节通过实验对方法进行验证和评估; 第 5 节进行讨论; 第 6 节对本文工作做总结, 并给出未来工作思路。

2 相关工作

同源性分析是进行攻击溯源的重要手段, 尤其在现实的网络安全业务场景中, 对于同源分析有强烈的实战需求。如果能在发现攻击后及时判断其攻击来源, 就可以采取针对性的应急响应和防护措施, 并提高对不同攻击源的认知, 扫除安全检测的盲区, 增强防御能力, 并对攻击者形成威慑。

根据线索类型不同, 现有的相关研究可以划分为恶意样本同源性分析和网络行为同源性分析。

2.1 恶意样本同源性分析

恶意样本同源性分析由于样本库作为威胁情报的重要组成部分容易获得、相关基础深厚等原因, 已经得到了广泛研究, 在学术界和产业界都形成了较为成熟的研究成果。恶意样本同源性分析具体可以划分为静态同源分析和动态同源分析。

在静态同源分析中, 研究人员从代码结构入手提取特征, 代码结构能够体现样本的语义逻辑和编写习惯, 因此来自同一作者的代码或存在函数复用的代码具有一定的同源特性。基于此, 样本的静态序列特征和图特征可以用于同源性分析[17]。静态序列特征包括 API 调用序列、函数调用序列等[18-20], 谭等人提取恶意样本的动态 API 序列特征和静态字节熵特征作为混合特征, 实现了恶意软件家族分类[21]。但 API 调用具有一定的通用性, 在表征恶意样本个性化特征方面存在短板。控制流图(Control Flow Graph, CFG)、函数调用图(Function Call Graph, FCG)、数据流图(Data Flow Graph, DFG)都以图的形式表示代码的执行过程, 控制流图表征代码块之间的跳转关系, 函数调用图表征代码中函数的调用关系, 而数据流图表征数据在样本执行中的逻辑流向。Alrabaee 等人[22]提出 BinGlod, 从样本中提取 CFG 并进一步提取 DFG, 构成语义特征流图, 利用最大公共子图算法进行了一系列语义特征对比工作。Suarez 等人[23]提取样本 CFG 以计算恶意样本家族指纹, 通过相似度计算进行同源性分析。Huang 等人[24]采用 CFG 最长路径比对的方法确定代码函数间相似性, 并进行相似性得分的计算。此外, 代码的抽象语法树也能够表示代码结构, Caliskan 等人[25]以代码的抽象语法树作为特征, 实现了对代码作者的识别。

样本编写者的知识水平、经验以及所使用的工具会对其编写习惯和风格产生潜在的影响, 包括代码布局、用词习惯等, 使其编写的代码存在相似性。Nataraj 等人[26]首次将恶意代码二进制文件转化为灰度图, 借助图像算法分析灰度图像纹理, 实现恶意代码分类。将代码与自然语言处理方法相结合也是一个研究热点, Kothari 等人[27]从代码中提取词汇, 并使用 n-gram[28]进行处理; Burrows 等人[29]将代码转化为字节集, 同样借助了 n-gram, 以实现作者身份的识别。在代码风格的研究上, Caliskan 等人[25]通过代码风格对源代码进行作者归属分析, 提取并构建了代码风格特征集。

在动态同源分析中, 通过在可控环境中监测恶意样本执行时的动态行为进行分析, 包括 API 调用

序列、系统调用参数等, 可以应对代码混淆、运行时打包技术[30]等导致的静态特征难以分析情况[31-33]。

虽然恶意样本同源性分析研究成果显著, 但在实际的应用场景中还存在一些问题。具体来说, 样本同源性分析更多地关注在恶意样本家族的同源, 而不是使用样本的攻击组织之间的同源。不同的攻击组织可能使用相同家族的样本, 随着黑色产业链条的商业化发展, 这种情况更加普遍, 影响了攻击同源性分析结果的准确性。此外, 已经出现了针对代码作者身份同源的攻击研究, 借助对抗机器学习模范编码风格, 导致错误的恶意样本同源结果[34]。因此, 安全研究人员在未来更加需要关注透露出攻击者行为模式的网络行为的同源性分析。

2.2 网络行为同源性分析

网络行为同源性分析方面的研究成果相对较少, 主要方法包括通过网络节点、数据包标记等[35-36]还原攻击路径、追溯攻击源头, 或利用通信链路上的网络行为特征进行分析, 包括 IP、URL、数据包、身份、地域、哈希值、时间戳几个维度的特征[37]。吴等人[38]提出了基于电子指纹的网络攻击溯源技术, 通过提取并组合数据流中的交互信息、时间戳等生成电子指纹并嵌入数据包, 从而分析攻击路径和进行攻击溯源。王建华等人[39]针对工控系统, 利用工控协议功能码的粗粒度统计特征和细粒度序列特征来量化攻击行为, 提取特征并构建模型, 进行同源分析。而王跃达等人[40]通过 HTTP 流量特征码匹配进行 Webshell 的检测与溯源。

对于学术界, 网络行为同源性分析领域缺少权威的数据集供学者研究; 而对于产业界, 由于安全监测设备每天都会产生海量告警, 这些告警具有如下特点:

(1) 同一次行动中发生时间不同但内容同质重复的告警多, 仅根据被攻击 IP、较短时间窗口和少量威胁情报标签进行简单归并和分类;

(2) 针对不同被攻击 IP 的同源攻击通常被划分到不同的类别, 对同源分析有负面影响;

(3) 误报多, 由于告警规则缺少灵活性, 导致大量的正常行为被命中, 从而产生误报。

现实的情况是, 各类机构实体部署了大量的安全防御产品和各类解决方案, 但由于上述原因, 安全分析团队淹没在海量告警中, 难以对己方遭受到的攻击有全面的认知, 特别是难以看见实际的攻击源, 从而影响应急响应的防御工作。

此外, 现有的攻击同源性分析过于依赖威胁情报, 但以 APT 组织为首的大量攻击者越来越注重身份的隐藏, 会最大程度避免直接在恶意代码、恶意域名注册记录等情报线索中遗留真实的身份信息或固定的痕迹信息, 且可能存在刻意伪装的虚假线索, 因此威胁狩猎过程难以通过威胁情报标签的简单关联确认同源。

相关工作描述及缺陷分析汇总如表 1。

本文围绕当前研究欠缺网络行为同源性分析的现状, 通过分析网络行为挖掘攻击者在攻击过程中难以被刻意伪造的潜在的行为模式, 从而不依赖先验威胁情报标签对攻击行为进行关联和同源分析。相比现有的相关工作, 本文工作具备以下几个方面的优势:

表 1 攻击同源相关工作对比

Table 1 Comparison of related works in cyberattack grouping

类别			技术描述	相关工作	存在缺陷
学术界	恶意样本同源性分析	静态同源分析	通过代码结构、函数调用序列、控制流图、代码风格等分析同源性	[18-29]	更关注恶意样本家族同源, 易通过借用、更换恶意工具逃逸同源分析
		动态同源分析	利用动态行为监测、API 调用序列、系统调用参数等分析同源性	[30-33]	
	网络行为同源性分析		数据包标记	[35-36]	针对特定场景, 欠缺普适性
			电子指纹嵌入	[38]	
			通信链路特征码	[37, 39-40]	
产业界	告警分析		分析海量告警进行一次攻击的关联发现	-	告警数据量大, 更关注攻击发现
	威胁情报		通过威胁情报标签关联做出同源判断	-	依赖于打了标签的已知特征, 可能存在特征伪装和同特征造成的混淆

(1) 在网络行为同源性分析领域开展研究, 补充当前研究空缺;

(2) 通过多维细粒度数据处理和智能算法分析来消化海量数据, 提高分析效率和数据有效利用率;

(3) 网络行为相比恶意样本、威胁情报线索, 不易被攻击者刻意伪装, 更能体现同一攻击者、攻击组

织的同源性。

3 网络行为同源分析方法研究

网络行为是攻击者在攻击实施过程中必然遗留的行为痕迹, 按检测结果可分为已被直接检测出的攻击告警和未被直接检测出其他相关行为。其中, 攻击告警直接表征着本次攻击行为的最鲜明特点, 在告警发生的同时未被检测出的同 IP 网络行为也包含着与攻击相关的行为模式, 如攻击者在发送含攻击载荷的数据包之前通常会有少量其他具有特定目的数据包。将攻击告警和告警发生时的较小时间窗口内的相关行为载荷关联起来形成的攻击行为中蕴含着攻击者更具个性化的攻击设施、技术手段、行为模式等特征, 因此, 利用网络行为进行攻击溯源研究是至关重要的。但由于权威数据集欠缺、有效数据难获取等现状, 目前的研究成果还不足以应用于现实中的行为同源分析场景。本文针对当前研究存在的问题与缺陷, 提出了一个基于网络行为的攻击同源分析框架, 通过多维度的攻击行为分析来抽取攻击者的深度行为模式, 并利用智能算法实现了攻击行为的聚类, 达到同源分析的目的。

3.1 框架设计

本文设计了一个基于网络行为的攻击同源性分析方法, 框架设计如图 1 所示。本文以从实验室采集的现网流量作为数据源, 首先做预处理, 以不同的 IP 进行告警划分, 并设计规则将各条告警识别为 5 个攻击阶段, 以便刻画攻击者在不同的攻击阶段中表现出的不同行为特征。随后, 从多个维度进行特征分析, 捕捉攻击者在攻击目标倾向、攻击设施利用、技术手段、行为习惯、个性化特征等方面的行为模式, 构建特征矩阵。为了评估不同 IP 间行为模式的相似程度, 并最大可能地保留攻击阶段与相应特征维度所体现的攻击者行为模式, 本文设计了特征矩阵相似度计算方法。最后, 利用该相似度数值, 借助社区发现算法将各 IP 划分到攻击组织社区, 实现攻击组织的同源分析。

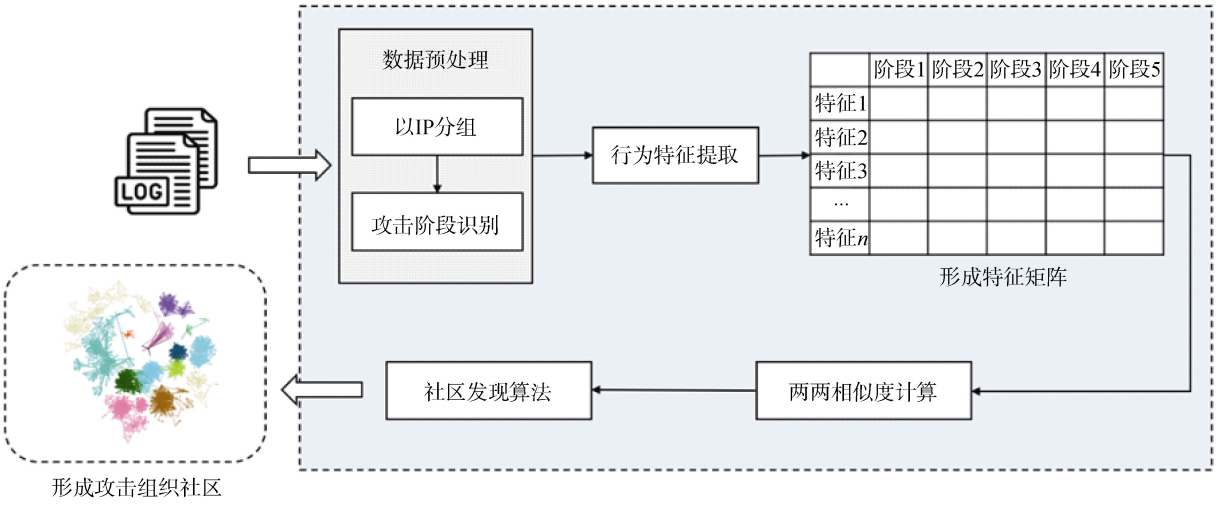


图 1 网络行为同源分析方法框架

Figure 1 Framework of network behavior-based attack attribution analysis method

3.2 数据预处理

由于告警的产生来源于大量攻击检测规则, 在检测过程中存在多个规则表征同一攻击行为的情况, 如果直接基于这些攻击特征进行计算, 将会产生大量噪声, 并且导致各阶段的攻击行为特征混杂, 无法正确表征攻击者的行为模式。为了将分散的攻击特征收敛到与攻击者行为模式更相关的维度, 本文没有从所有的攻击检测规则中直接进行特征提取, 而是对每一次攻击进行了攻击阶段的识别和划分。攻击阶段参考杀伤链模型[41], 划分为侦查、入侵、命令控制、横向移动、目的执行 5 个阶段, 其中, 由于武器化阶段的网络行为不在本文流量能采集到的范围而被去除, 投递阶段由于本文不进行样本检测而被去除。一般来说, APT 由于具有明确的攻击意图和成熟的技战术, 会涉及多个攻击阶段, 而普通网络攻击可能只涉及 1 或 2 个攻击阶段, 我们对不同的攻击阶段提取相同维度的特征形成特征矩阵, 因此该框架对于 APT 和普通攻击是通用的。

为了识别攻击阶段, 本文对富化后的告警数据进行了两层分类分析, 第一层分类是根据告警的规则库和网络载荷特点, 基于我们的经验, 结合 ATT&CK 模型[42], 将网络行为数据分为不同的威胁手法类型, 并进行了统计和归并, 共得到 28 种攻击行为类型, 如将木马通信载荷里命中的 IP 特征映射

到远控木马通信类, 而爬虫在 GET 请求中的 IP 特征映射到目录探测类; 第二层分类是根据攻击者在不同阶段的常用手法, 将这 28 种行为类型映射到相应的攻击阶段, 具体如表 2 所示。

表 2 攻击阶段映射规则

Table 2 Rules of attack phase definition

攻击阶段	威胁类型
侦查	目录探测
	端口和服务扫描
	漏洞扫描
入侵	代码执行
	非授权访问/权限绕过
	命令代码执行
	Webshell 上传
	SQL 注入
	口令暴力猜解
	跨站脚本攻击
	恶意代码投递
	文件读取
	溢出攻击
	伪装 URL 跳转
	其他漏洞利用行为
命令控制	黑市工具通信
	Webshell 通信
	APT 武器通信
	网络蠕虫通信
	僵尸网络通信
	远控木马通信
	流氓广告插件通信
	其他恶意软件通信
	后门程序通信
横向移动	对配置不当的利用行为
	横向移动工具通信
目的执行	文件下载
	信息泄漏

3.3 特征提取

攻击者的攻击过程由 5 个阶段组成, 分别为侦察阶段、入侵阶段、命令控制阶段、横向移动阶段和目的执行阶段。

从同源分析的角度来说, 我们分析同一组织的攻击行为特性。对于属于同一个组织的攻击者, 其所处的地理位置以及惯用的设施资源具有一定的稳定性; 在攻击目标方面, 同一组织往往会对某个行业或某类特定目标具有倾向性; 此外, 由于执行攻击行动的人会有一些个人风格和行为习惯, 其在攻击行为上也可能局限于某种行为模式, 体现出符合一定规律的特点。

本文总结了团队在合作方现网流量中追踪 APT 组织和人工进行网络行为同源分析的实战经验, 参考 2008 年 10 月 1 日以来至 2021 年 9 月 23 日的 554 篇 APT 组织分析报告[43], 从中剔除了恶意代码相关的特征后设计了基于网络行为对攻击组织进行刻画和同源分析的特征。在每个攻击阶段中, 我们提取了 4 个特征类别, 包括攻击目标、攻击设施属性、活动规律、个性化特点。其中攻击目标包含受害者 IP、受害者域名 2 个特征; 攻击设施属性包含攻击者 IP 的 B 段、攻击者 IP 的 C 段、攻击 IP 所在国家、攻击 IP 所在城市 4 个特征; 活动规律包含威胁手法、请求方法、User-Agent(UA)信息、代理服务器使用、X-Forwarded-For(XFF)、活动时间间隔 6 个特征; 个性化特点包括活动时间段、特殊位置字符串 2 个特征, 具体如表 3 所示。

表 3 攻击行为特征提取

Table 3 Features extraction of attack behavior

攻击组织属性类型	具体属性	特征
固有属性(静态)	攻击设施属性	攻击者 IP 的 B 段
		攻击者 IP 的 C 段
		攻击 IP 所在国家
		攻击 IP 所在城市
	攻击目标	受害者 IP
		受害者域名
行为属性(动态)	活动规律	威胁手法
		请求方法
		UA
		代理服务器使用
		XFF
		活动时间间隔
	个性化特点	活动时间段
		特殊位置字符串

根据上述特征设计, 为每个 IP 提取一个 5**n* 的特征矩阵, 其中 5 表示 5 个攻击阶段, *n* 表示每个阶段由代表 4 类特征类别的的 *n* 维向量组成。

对于攻击目标类别, 通常来说, 攻击者的攻击行为具有一定的针对性和偏好性, 同一攻击组织很可能对某些特定目标具有更高的兴趣, 因此在相对较集中的时间窗口内, 对相似目标的攻击行为更有可能来自同一组织, 因此我们提取了受害者 IP、受害者域名 2 个特征, 我们对每一个 IP、域名值赋予一个独立的编号作为特征值。

对于攻击设施属性类别, 为避免在执行恶意操作时暴露真实 IP, 攻击者通常同时或轮流使用从同一代理供应商购买的一个代理或大量代理 IP, 因此攻击者很可能在同一 C 段和 B 段中使用多个代理 IP, 也导致攻击 IP 属于同一个城市或国家。我们也对每

一个 B 段、C 段、城市、国家赋予一个独立的编号作为特征值。

对于活动规律类别, 考虑到同一攻击组织所具备的知识和技能, 以及人类活动会受习惯所影响的特性, 其在开展攻击活动时会产生惯性行为, 使用相同或相似的威胁手法, 构造模式或相似的攻击语句。UA 是 HTTP 协议中的一部分, 属于头域的组成部分, 其中包含了攻击者所使用的浏览器类型、操作系统、浏览器内核等信息的标识, 同一攻击者也很可能向 UA 中嵌入相似的恶意代码或使用伪造的相似 UA。在 HTTP 协议头域中用 x-via 字段记录代理服务器使用情况。XFF 字段同样存在于 HTTP 协议头域中, 用于获取访问用户的真实 IP。同一攻击组织在以上攻击信息中都具有固定性和相似性, 这些字段虽然不能直接指征同源的攻击行为, 但具有较高的旁证价值, 都能够成为我们对攻击组织同源性分析的线索, 因此我们提取了威胁手法、UA、代理、XFF 这些特征。上述特征不同于攻击目标、攻击设施属性类别等特征, 每个攻击来源可能使用多种威胁手法、代理、UA 等, 不能使用一维的编号作为特征值。因此, 我们同样给每一个值赋予一个独立的编号, 采用 One-Hot 方法构建多维向量, 作为特征。而对于威胁手法特征, 由于种类数量特别巨大, 而每个攻击者使用的手法有限, 因此采取了不同的处理方法, 进行编号序列化转换, 对各 IP 使用的威胁手法对应的编号组成一个序列, 以使用手法最多的数量作为序列长度, 保证特征维数统一, 序列不足该长度的用特殊数字补位。在提取攻击时间间隔时, 我们设置了 3 个级别, 分别为毫秒级、秒级和分钟级, 毫秒级表示攻击活动非常频繁, 不同攻击发动的时间仅间隔几毫秒, 通常见于探测活动; 秒级表示攻击活动比较频繁, 不同攻击以几秒的时间间隔发动, 可能是扫描探测, 也可能是漏洞利用的尝试行为; 分钟级表示攻击活动的时间间隔为几分钟甚至更长的时间, 可能是慢速窃取行为或一些需要人工交互的操作。可见, 上述不同级别的时间间隔与攻击者的目的有关, 也能在一定程度上指征攻击者的手法。

对于个性化特点类别, 主要考虑攻击者的工作时间段和特殊位置字符串。不同的攻击组织在攻击时间上也具有规律性, 比如有的组织作息规律, 攻击时间集中在早上 9 点到下午 5 点, 有些组织习惯中午开始活动, 攻击行为持续到深夜时分, 有的组织则习惯在凌晨时段活动, 同样的, 相同的攻击组织在攻击间隔上也具有规律性和稳定性。因此我们在提取攻击时间的特征时, 以小时为单位将一天分为 24 段, 用 24 个维度来表示, 记录每一段对应的攻击数量所占全天总攻击数量的比例。此外, 最能体现同一组织或攻击者个性化特点的是特殊位置字符串, 因为特殊位置字符串是依赖个人具体个性构造的, 具有极高的标示性, 比如攻击者自定义的参数值、参数名、文件名等字符串。我们在提取该特征时使用了 Word2vec 模型[44], Word2vec 模型由神经网络组成用来训练以重新建构词文本, 训练完成之后, Word2vec 模型可用来映射每个词到一个向量, 可用来表示词对词之间的关系。我们首先使用丰富的 wiki 语料库作为训练集训练 Word2vec 模型, 然后利用训练好的 Word2vec 模型将特殊位置字符串转化为维数为 100 的向量形式, 作为我们所需要提取的特征。

每个恶意 IP 会发起一到多条攻击活动, 本文根据 5 个阶段的活动分别进行上述所有特征维度的提取, 最终组合得到一个特征矩阵来刻画行为模式。特征矩阵横向上有 5 列, 对应 5 个攻击阶段; 纵向上行数不定, 且可根据新攻击手法、攻击资产的发现进行扩展。为避免特征矩阵过于稀疏导致后续计算不准确, 将维数较高的特征进行了两种方式的处理, 具体来说, 对维度超过 1,000 的特征采用主成分分析法 (Principal Component Analysis, PCA)[45]进行降维, 统一降到 200 维, 这些特征包括受害者 IP、受害者域名、UA、XFF; 对威胁手法特征进行编号序列化转换, 维数为 117。最终的特征矩阵维度为 5*1110, 各特征维度具体如表 4 所示。

3.4 行为相似度分析

接下来, 利用特征矩阵来计算各恶意 IP 间行为模式的相似程度。根据不同维度特征形式的不同, 采用不同的算法进行相似度计算: 对于 IP 固有的地理位置、IP 段特征, 由于不受攻击阶段影响, 因此只判断是否相同, 并分别用 0 或 1 进行表示; 对于不同攻击阶段表现不同的特征维度, 则每一类特征都是特征矩阵中的子矩阵, 借助矩阵相似度算法分别进行计算, 具体来说, 对于威胁手法序列采用杰卡德相似指数[46]进行计算, 对于其他特征采用余弦相似度算法[47]进行计算。

两两 IP 间的相似度 *Similarity* 计算如式(1)所示。

$$Similarity = \sum v_i w_i \tag{1}$$

其中, v 表示 IP 间各特征指标的相似度数值, w 表示各特征指标的权重, 根据各类特征对于攻击者行为独特性表征能力, 我们为个性化特征类别赋予了相对更高的权重。

表 4 攻击行为特征矩阵

Table 4 Feature matrix of attack behavior

	侦查	入侵	命令控制	横向渗透	目的执行
城市	1	1	1	1	1
国家	1	1	1	1	1
攻击 IP B 段	1	1	1	1	1
攻击 IP C 段	1	1	1	1	1
受害者 IP	200	200	200	200	200
受害者域名	200	200	200	200	200
威胁手法	117	117	117	117	117
请求方法	7	7	7	7	7
UA	200	200	200	200	200
代理服务器	55	55	55	55	55
XFF	200	200	200	200	200
活动时间间隔	3	3	3	3	3
活动时间段	24	24	24	24	24
特殊位置字符串	100	100	100	100	100

接下来, 本文以 IP 为节点, 以 IP 间行为模式相似度作为边权重构建 IP 行为网络图, 即作为判断 IP 间关联程度进而识别攻击组织的依据。

3.5 基于社区发现算法的攻击组织划分

出于判断不同 IP 所属攻击组织关系的目的以及缺少权威训练数据集的现状, 本文采用无监督的聚类算法来分析 IP 行为网络图。在类型众多的聚类算法中, 基于密度的聚类算法如 DBSCAN、基于树形结构的聚类算法如孤立森林等都更加关注节点的属性, 旨在找到属性相似的节点, 从而忽略了节点之间的联系, 而社区发现算法则依据节点间的关联进行社区划分, 侧重于找到网络中联系紧密的部分, 这更符合本文的需求和数据特点。由于攻击组织数量未知、攻击 IP 数量庞大, 本文需要选择一种高效的社区检测算法来从大型加权网络图中提取社区。

Louvain 算法是一种基于聚类的社区划分算法, 通过模块化的增益迭代地合并节点, 能够快速有效地辨别有层次的社区结构从而对大型网络进行社区划分, 具有快速、准确的特点, 被认为是性能最好的网络或图的发现算法之一[48], 因此本文选用 Louvain 算法。在下文中也通过实验证明了该算法的优越性。

在 Louvain 算法中有两个主要参数, 模块度(Modularity, 记为 Q)以及模块度增量(Delta Modularity, 记为 ΔQ)。其中模块度 Q 能够描述划分的社区内部节点的紧密程度, 是衡量发现算法结果质量的重要参数, 模块度越大表明划分效果越好。模块度函数定义如下:

$$Q=\frac{1}{2m}\sum_{i,j=1}^{n}[A_{ij}-\frac{k_ik_j}{2m}]\delta(c_ic_j) \tag{2}$$

式中, A_{ij} 为任意节点 i 与 j 之间的边权重, k_i 与 k_j 分别为所有节点与任意节点 i 和 j 之间的边权重, m 表示网络中的边数总和, $\frac{k_ik_j}{2m}$ 表示平均边权重, $A_{ij}-\frac{k_ik_j}{2m}$ 表示该网络的真实结构和随机组合时的预期结构之间的差。c 为社区, c_i 和 c_j 分别表示节点 i 和 j 所在社区, 函数 δ 表示属于同一社区, 当 δ=1 时, 表示节点 i 和 j 属于同一社区, 当 δ=0 时, 表示节点 i 和 j 不属于同一社区。

Louvain 算法在社区划分过程中, 每当一个新的节点加入到某个社区后需要重新计算社区的模块度, 新节点加入后社区模块度的变化量用模块度增量 ΔQ 来衡量, 计算公式如下:

$$\Delta Q=\left[\frac{\sum in+k_{i,in}}{2m}-\left(\frac{\sum tot+k_i}{2m}\right)^2\right]-\left[\frac{\sum in}{2m}-\left(\frac{\sum tot}{2m}\right)^2-\left(\frac{k_i}{2m}\right)^2\right] \tag{3}$$

其中, $\sum in$ 表示社区内部边权重之和, $\sum tot$ 表示所有与社区内部相连的边权重之和, $k_{i,in}$ 表示节点 i 加入到社区 c 中时的权重之和。当 Q 的值不在变化时, 此时停止计算, 说明所有的节点都被分组成了一个巨型聚类或者已有的类无法进一步合并。

在 Louvain 算法中, 解析度这一参数可以灵活控制社区划分的过程, Lambiotte[49]认为使用模块度为质量函数的社区算法容易对特定结构的模型产生偏差, 因此加入了解析度来灵活控制社区划分的数量以及规模。

4 实验与评估

在本节, 我们基于真实数据集开展攻击同源实验, 验证本文方法效果, 并进行案例分析与评估。

4.1 数据集

网络安全领域的数据集普遍存在着理论和实践有差距的问题, 可能的原因在于欠缺部分特定领域需求、权威可信数据的共享机制不完善等[50]。本文没有采用 ISCX-IDS 2012、CIC-IDS 2017 等公开的实

验性攻防数据集, 而是采集了实验室与合作伙伴在日常工作中产生的实际流量, 经匿名化后生成了数据集。不采用实验性数据集原因主要有以下三点:

(1) 实验室生成的模拟数据很难覆盖真正的攻防场景, 特别是使用了多攻击设施、涉及多攻击阶段的行动;

(2) 实验性数据集缺乏真实的攻击源数据, 如国家和城市等, 也缺乏详细的能指征不同攻击组织习惯的攻击手法数据, 如是否使用代理、转发的 IP、攻击的详细载荷等, 基本失去了对攻击行为组织性的表征能力;

(3) 现成的实验数据集均已提取好攻击特征字段, 这些固定的特征字段无法满足我们提取攻击组织属性特征的需求。

针对上述实验性数据集的不足, 我们采集了时间跨度为 30 天的实际流量设计和生成了自己的数据集, 并且经过了检测和富化, 形成了攻击行为日志数据:

(1) 数据采集。基于实验室的入侵检测设备和覆盖攻击链各过程的检测规则, 生成入侵检测告警;

(2) 告警过滤。将入侵检测设备未归并的重复告警日志和依据经验能明显区分出的误告警筛除;

(3) 数据富化。基于网络基础情报关联补充攻击 IP 的城市、国家等相关属性, 基于告警 IP 的网络审计日志关联补充告警 IP 在同时刻的其他行为属性, 如 UA、所使用 CDN 节点、代理 IP 等;

(4) 生成攻击行为数据集。

该数据集中包含了基于先验情报的攻击行为来源标签, 由入侵检测系统根据人工核验过的威胁情报关联打标生成。在我们的实验过程中去掉了这些来自威胁情报的共 46 个先验标签, 仅用于和实验结果进行对比, 评估结果的准确度。

最终得到的数据集包含 5,104 个 IP 共 114,845 条告警, 每条告警包含 11 个可分析字段, 具体如表 5 所示。

4.2 结果分析

我们根据 IP 间行为相似性构建了 IP 行为网络图, IP 行为网络图 G 由 IP 节点和 IP 节点之间的边构成, 即 $G=\{V,E\}$, 其中 V 表示节点的集合, E 表示边的集合, 边的权重值由两节点间的行为相似性决定, 即 $A_{ij}=sim(i,j),i,j\in V$。然后我们使用 Louvain 社区划分算法对 IP 行为网络图进行攻击社区划分, 并对实验结果进行分析。

4.2.1 攻击社区划分结果

通过相似性计算, 我们得到了 IP 节点间的边的权重值, 其中权重最大值为 19.0, 最小值为–2.196, 平均值为 4.688。需要注意的是, 当前 IP 节点两两之间都会存在一条边, 这使得 IP 行为网络图的体积过于庞大且不利于社区划分, 因此我们通过设置边权重阈值来确定某条边是否载入 IP 行为网络图中, 具体如下公式所示:

$$G.add\_edge(i,j)=\begin{cases}True, A_{ij}>threshold\\ False, A_{ij}<threshold\end{cases} \tag{4}$$

Louvain 算法的默认解析度为 1.0, 当解析度为 1.0 时, 我们测试不同阈值下的社区划分效果, 如图 2 所示, 从图中我们可以看出当阈值为 14 时, 对应的模块度最高为 0.7808, 社区划分效果最好, 因此我们选用边权重阈值为 14 来进行网络图构建。至此, IP 行为网络图已经构建完成, IP 行为网络图结构如表 6 所示。

表 5 数据集说明

Table 5 Datasets

字段	说明
攻击发生时间	包含日期和时间, 格式为“y-m-d h:M:s”
攻击 IP	攻击者的 IP 地址
攻击 IP 地理信息	经数据富化获得, 包括攻击 IP 所在的国家、城市
请求	攻击 IP 向受害 IP 发送的请求头及请求体
响应	受害 IP 向攻击 IP 返回的响应头及响应体
URI	攻击者发送请求的路径
XFF 代理	XFF 代理信息
载荷内容	攻击载荷内容
威胁手法	基于载荷分类获得
受害 IP	被攻击的 IP 地址
受害域名	被攻击的域名

表 6 IP 行为网络图结构

Table 6 Structure of IP behavior network graph

节点数量	边数量	最小边权重	最大边权重	平均边权重
4334	139653	14.0001	19.0	15.4828

我们使用 Louvain 社区划分算法对 IP 行为网络图进行社区划分, 我们测试了不同解析度下的社区划分效果, 如图 3 所示, 从图中看出解析度为 1.0、1.5、2.0 时模块度最高为 0.7808, 社区划分效果最好, 因此我们选用解析度为 1.0 来进行社区划分。

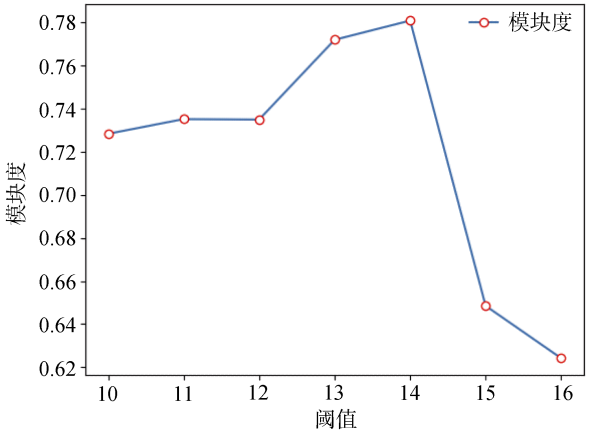


图 2　不同阈值下模块度变化

Figure 2 Variation of modularity under different thresholds

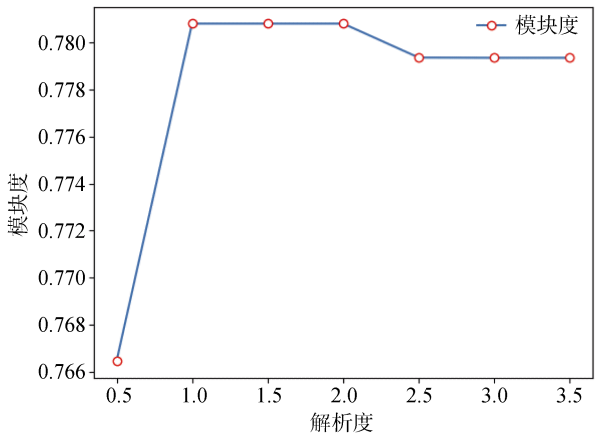


图 3　不同解析度下模块度变化

Figure 3　Variation of modularity under different resolution

我们对社区划分结果进行了进一步处理，为排除离散点对于社区划分效果的影响，我们筛除了包含节点数小于总节点数百分之一的社区，主要原因一是实验数据采集的时间窗口有限，很多具有组织性的攻击行为时间跨度很大，超出了采集时间范围因而不够完整；二是根据对攻击 IP 的统计，入侵检测数据采集的范围内存在“长尾效应”，40%以上的攻击行为是散发的，攻击 IP 可能只出现一次且仅有一个威胁情报标签，无法匹配到能够指征更多行为特点的同一时间网络行为数据；三是过滤后的数据集可能仍存在少部分误告警。

在排除了上述对数据的影响后，最终留下 3,081 个 IP，社区划分可视化效果图如图 4 所示，共有 14 个社区，不同的颜色代表不同的社区，从图中可以看出社区间划分清晰，划分效果良好。

4.2.2　算法对比

在本节，将 Louvain 算法与社区发现领域另外两

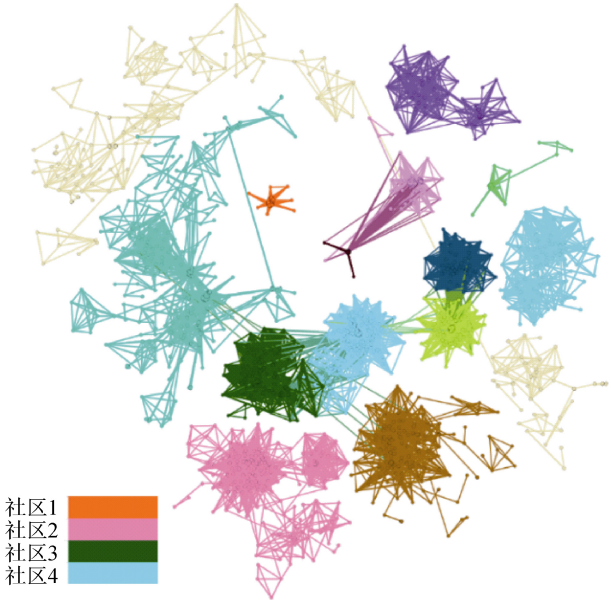


图 4　社区可视化效果图

Figure 4　Community visualization

个常用的算法——Infomap 和 Girvan-Newman(GN)算法进行了对比。Infomap 算法采用随机游走，通过在密集社区内的随机游走来识别并划分社区。GN 算法是基于分裂的社区检测算法，基于低相似性删除节点之间的边从而将社区彼此分离。

本文对三种算法分别进行了实验，三种算法最优参数下的社区划分效果如表 7 所示。

表 7　算法效果对比

Table 7　Comparison of results of three algorithms

算法类型	最优模块度
Louvain	0.7808
Infomap	0.6566
GN	0.6565

Louvain 算法取得了更高的模块度，证明其社区划分效果更好。经过分析，GN 算法在机制上不能判断算法终止位置，且可能存在重复计算导致时间复杂度更高；而 Infomap 算法与 Louvain 算法都以优化目标函数为驱动，前者倾向于划分小而规模平均的社区，后者更能发现社区间的差异性，图规模越大则 Louvain 算法的优势更明显。因此，本文选取 Louvain 算法可以起到更好的效果，而且在应对后续更大规模的攻击组织分析上具有更好的适应性。

4.2.3　案例分析

根据图 4 的社区聚类结果，可以看出在 5,104 个 IP 的共 114,845 条告警日志中，去除了离散点后共有 14 个明显的社区，这些社区大致分为三个类型：完

全独立的社区、内部包含多个分簇的社区、距离很近的两个社区。下面从每个类型中选取一个典型的案例进行分析。

(1) 案例 1: 社区 1, 与其他社区距离都较远, 相对独立且致密。

我们对该社区的数据进行了分析, 发现虽然该社区在图中占位较小, 但实际上是最致密的社区之一, 共包含 361 个攻击 IP。这些 IP 共分布在 7 个 C 段, 图形中正好有 7 个紧凑的分支, 正好能够与这 7 个C段对应。这7个C段的IP活动时间均集中于2021年 6 月 18 日至 6 月 30 日, 攻击行为间隔分布均匀, 最大间隔不超过 3 秒。涉及到攻击目标仅有 2 个单位, 但攻击行为中涉及到的特点覆盖了表 1 中侦查、入侵和命令控制三个阶段的所有威胁类型。我们经过对攻击和响应载荷数据的分析研判, 结合攻击 IP 池在威胁情报中的关联, 并向实验室合作方进行确认, 得出结论认为该社区的攻击行为是国内某安全公司对实验室合作伙伴进行的渗透性安全测试, 主要是漏洞探测和木马感染探测, 因而仅包含了侦查、入侵和命令控制阶段, 缺乏横向移动和窃取行为。在数据集的先验标签中, 该社区的所有攻击行为的来源都被标记为“Axiom”标签, 这与我们的实验结果是吻合的, 因为 Axiom 正是一个可扩展的渗透测试框架, 被多家安全公司广泛用于各类安全性渗透测试工作。

(2) 案例 2: 社区 2, 有三个明显的分簇, 但隶属于同一个社区。

我们对该社区的数据进行了分析, 发现该社区共有 196 个不同的 IP, 但与案例 1 中的致密情况相反, 这 196 个 IP 分散于 119 个不同的网段(89 个 B 段中的 119 个 C 段), 在时间分布上则分散于整个数据采集时间区间。从攻击行为上来看则类型集中, 有 94% 的攻击行为落在命令控制区间, 在入侵检测出来的共 31 种标签中, 属于中国菜刀、蚁剑、跨平台版中国菜刀 CKnife 等典型的僵尸网络通信行为的通信共 14657 条, 占该社区产生的攻击行为日志总数的 81.2%, 且受攻击目标不存在特殊的分布规律。分析人员手动选取了一些载荷, 发现大多数载荷属于心跳, 且属于不同子簇的心跳载荷不一致, 这些不同的载荷在实验前的载荷分类中被分入了命令控制阶段中不同的威胁类型, 这可能是导致该社区包含有几个子簇的原因。综上分析, 可见该社区应当隶属于一个规模较大且活跃的僵尸网络, 且该社区的 IP 应当处于僵尸网络的中间层位置, 即这些 IP 不是最上游的控制端, 同时又控制有下游的一些“肉鸡”节点。在数据集的先验标签中, 该社区的所有攻击行为都被标记为“soft cell”和“botnet”标签, 该聚类结果符合先验的标签知识。

(3) 案例 3: 社区 3 和社区 4, 两个社区距离紧密, 且有交叉。

在社区划分的结果中, 上述两个社区引起了我们的注意: 两簇 IP 距离很近, 但被划分到了不同的社区。这两个社区分别包含 298 和 143 个攻击 IP, 经过验证, 社区 4 的主要攻击行为均为开源的商业攻击框架和已经被曝光的 Office CVE 漏洞, 而社区 3 的主要攻击行为则是垃圾邮件、口令猜解等相对低端一些的类型。两个社区的攻击设施类型中均有 VPS(Virtual Private Server, 虚拟专用服务器), 并且其中出现了一个共同的供应商: 阿里云。社区 4 的 VPS 主要来自 DigitalOcean 和阿里云, 社区 3 的攻击设施则比较杂乱, 除了少数租用来自 Linode、阿里云的 VPS 之外, 多数是来自运营商的 IDC 机房, 还有一些被标为家庭宽带、企业专线的设施。分析人员对上述两个社区的攻击设施进行了手动验证, 来自社区 3 的攻击设施缺乏统计上的共同点, 符合受控僵尸网络肉鸡的特性, 来自社区 4 的攻击设施则具有较特殊的共同点: 禁 Ping、网络空间测绘引擎(shodan 和 fofa)中无记录、大部分已完全关闭、探测到的存活 IP 使用相同版本的 SSH 组件, 扫描结果见图 5。

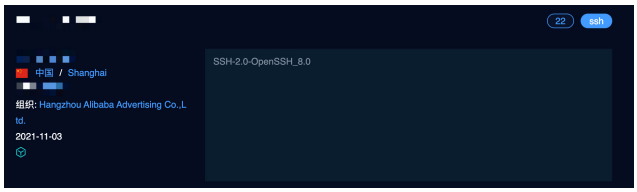


图 5 社区 4 存活节点的 SSH 组件版本

Figure 5 SSH version of existing nodes in Community 4

综上, 社区 3 和社区 4 从攻击设施、攻击行为模式和攻击目的上来看, 确实分属于两组不同的攻击者。在先验标签中, 社区 4 的 IP 既具有 Patchwork APT 组织的标签, 也具有僵尸网络的标签。我们经过深入分析, 认为即该社区不一定是 Patchwork APT 组织, 因为 Patchwork 组织本身是一个偏好使用各类开源工具、漏洞和组件的 APT 组织, 目前有的信息只能证明该社区可能使用了与 Patchwork 同样的开源工具, 是否确定是该组织还需要获取更多数据, 进行更长期的研究和分析工作。

4.3 实验评估

为了对实验结果进行验证、评估本文所提出的

网络行为攻击同源模型, 设计了三个评估指标: 准确率、召回率和社区同质率。我们的评估针对经过离散点筛除后的 IP 进行。

准确率 *Acc* 用于评估攻击 IP 是否能够准确划分到其所属的攻击组织。以模型运算后每个社区中 IP 所属攻击组织数量最多的作为该社区的组织标签, 计算正确划分的 IP 比例, 如式(5)所示:

$$Acc = \frac{m_{correct}}{M} \tag{5}$$

其中, $m_{correct}$ 表示组织划分正确的 IP 数量, 即实验结果中认为隶属于同一社区的 IP 在先验标签知识中也隶属于同一组织, M 表示 IP 总数。

召回率 *Recall* 用于评估各先验攻击组织的 IP 能否识别为该组织。同样以模型运算后每个社区中 IP 所属攻击组织数量最多的作为该社区的组织标签, 对每一个先验攻击组织单独计算召回率, 最后计算所有组织的综合召回率, 如式(6)所示。

$$Recall = \frac{1}{L}\sum_{i=1}^{L}\frac{k_{correct}}{K_i} \tag{6}$$

其中, L 表示先验攻击组织数量, K 表示各组织中 IP 数量, $k_{correct}$ 表示该组织中通过算法识别的攻击社区与先验攻击组织一致的 IP 数量, 通过求均值的方式计算综合召回率。

社区同质率 H 用于评估划分到每个攻击组织社区内的 IP 是否同源。以每个社区内最大同源 IP 数量计算该社区的最高同源率, 综合所有社区计算社区同质率, 如式(7)所示。

$$H = \frac{\sum_{i=1}^{N}\frac{\max(m_x)}{size_i}}{N} \tag{7}$$

其中, N 表示社区数量, *size* 表示社区大小, 即社区内的 IP 数量, m_x 表示该社区内属于同一组织 x 的 IP 数量, 而 *max*()表示取最大值。

上述 3 个评估指标的结果如表 8 所示。

表 8 评估结果

Table 8 Result of evaluation

评估指标	指标数值
准确率	0.9679
召回率	0.9354
社区同质率	0.9048

可见, 与先验标签对比, 本论文所提出算法的准确率和对攻击行为组织性的判断结果已经足够为分析人员提供可靠的同源分析参考。

5 讨论

同源分析是进行攻击溯源的重要手段, 尤其在实际的网络安全业务场景中, 对同源分析方法进行研究有重要的现实意义。本文提出的方法是基于真实的攻防数据集设计和验证的, 具有较高现实应用意义。

虽然本文的工作能够提供准确率达 96%的攻击行为同源分析, 但仍存在一定局限性。

(1) 本文的同源分析基础是富化后的告警数据, 因此对入侵检测结果的准确度有天然的依赖性。如果先期的入侵检测数据中误告警较多, 可能会导致本算法的社区划分结果中出现一定的离散点和错误。在本算法应用于现实的攻防场景时, 建议应用者和本文的数据准备工作保持一致, 首先对告警数据进行筛选和归并, 过滤掉误告警和没有实际追踪价值的告警信息, 从而保证方法的有效性。

(2) 目前本文只根据威胁情报、地理信息数据和告警产生同时刻的攻击 IP 上下文网络日志做了告警的情报富化, 数据中的场景仅局限于告警发生时的较小时间窗口, 这也是使得社区划分的准确率和同质性较高的原因。如果对时间窗口进行更多拓展, 引入范围更大的攻击 IP 上下文网络行为日志, 可能造成准确率的下降。

(3) 本文的方法不能代替同源分析领域的人工分析工作。攻击组织的行为模式对于防御方不总是可理解或可预测的, 深度同源分析和溯源仍需要大量分析人员参与, 并且为了使同源分析结论具备足够的置信度, 在一定场景下需要更主动地对攻击源 IP 进行探测。

6 结论

本文针对当前攻击溯源工作的薄弱点, 提出了一种基于网络行为的攻击同源分析方法。首先, 根据攻击在不同阶段的行为特点, 将各条攻击行为识别为侦察、入侵、命令控制、横向移动、以及目的执行 5 个攻击阶段, 从而刻画不同攻击阶段中的不同行为特征。然后, 从多个维度对来自各 IP 的攻击行为进行特征分析, 提取了攻击目标、攻击设施属性、活动规律、个性化特点 4 个类别共 14 个特征以抽象攻击行为模式, 构建特征矩阵。随后对特征矩阵进行了两两的相似性计算, 再以此为权值建立 IP 行为网络图, 借助社区发现算法划分攻击社区, 实现攻击组织同源。为了使本文的方法更具有现实价值, 本文未采用实验性的公开数据集, 而是基于真实攻防数

据采集和生成了数据集, 并基于该真实数据集设计特征和算法并进行实验验证, 准确率达到96%, 结果证明了方法的有效性。本文存在一定的局限性, 如依赖入侵检测告警的准确性、选取的攻击场景时间窗口较窄等, 未来工作将进一步扩大同源分析的基础场景, 不再仅针对告警数据进行简单富化, 而是围绕告警行为上下文的业务场景, 关联告警数据和攻击 IP、受攻击目标更多的网络行为数据, 如攻击 IP 的其他 DNS 解析行为、受到攻击的具体应用或协议等, 抽取更丰富的攻击源行为模式特征表示, 围绕更多的业务场景适用性开展同源分析工作。

致　谢　本论文实验中涉及的数据采集难度大、数据处理工序复杂、人工分析工作量大, 在此向给予指导的老师、提供帮助的同学和给本文提出建议的评审专家表示感谢。另外特别感谢国家网络空间威胁情报共享开放平台(CNTIC)和国家互联网应急中心(CNCERT)提供部分检测特征和供数据富化的网络基础信息。

参考文献

[1] 2020 年中国互联网网络安全报告.国家互联网应急中心. https://www.cert.org.cn/publish/main/upload/File/2020 Annual Report.pdf. Jun. 2021.

[2] 腾讯安全 2018 年高级持续性威胁(APT)研究报告. 腾讯安全. https://paper.seebug.org/781/Jan. 2019.

[3] 2020 全球高级持续性威胁(APT)研究报告. 360 安全大脑. https://www.360.cn/n/11878.html. Feb. 2021.

[4] Stone R. CenterTrack: An IP Overlay Network for Tracking DoS Floods[C]. *9th USENIX Security Symposium (USENIX Security 00)*, Denver, CO. July 2000. https://www.usenix.org/legacy/ publications/ library/proceedings/sec2000/full_papers/stone/stone. pdf.

[5] Savage S, Wetherall D, Karlin A, et al. Practical Network Support for IP Traceback[C]. *The conference on Applications*, *Technologies*, *Architectures*, *and Protocols for Computer Communication*, 2000: 295-306.

[6] Belenky A, Ansari N. IP Traceback with Deterministic Packet Marking[J]. *IEEE Communications Letters*, 2003, 7(4): 162-164.

[7] Alsaheel A, Nan Y, Ma S, et al. ATLAS: A Sequence-based Learning Approach for Attack Investigation[C]. *30th USENIX Security Symposium*. 2021: 3005-3022.

[8] Hassan W U, Guo S, Li D, et al. Nodoze: Combatting threat alert fatigue with automated provenance triage[C]. *Network and Distributed Systems Security Symposium*. 2019. https://www.ndss-symposium.org/wp-content/uploads/2019/02/ndss2019_03B-13_UlHassan_paper.pdf.

[9] Yu L, Ma S, Zhang Z, et al. ALchemist: Fusing Application and Audit Logs for Precise Attack Provenance without Instrumentation[C]. *Network and Distributed System Security Symposium* 2021. https://www.ndss-symposium.org/wp-content/uploads/ndss2021_7A -2_24445_paper.pdf.

[10] Ho G, Sharma A, Javed M, et al. Detecting Credential Spearphishing Attacks in Enterprise Settings[C]. *The 26th USENIX Conference on Security Symposium*, 2017: 469-485.

[11] APT attacks on industrial companies in 2020. Kaspersky. https://ics-cert.kaspersky.com/reports/2021/03/29/apt-attacks-on-industrial-companies-in-2020/ . Mar.2021.

[12] 2021 年上半年全球高级持续性威胁(APT)研究报告. https://cert.360.cn/report/detail?id=6c9a1b56e4ceb84a8ab9e96044429adc. Oct. 2021.

[13] APT attacks on industrial organizations in H1 2021. https://ics-cert.kaspersky.com/reports/2021/10/26/apt-attacks-on-industrial-organizations-in-h1-2021/. Oct.2021.

[14] Li V G, Dunn M, Pearce P, et al. Reading the Tea Leaves: A Comparative Analysis of Threat Intelligence[C]. *The 28th USENIX Conference on Security Symposium*, 2019: 851-867.

[15] Li Q, Yang Z M, Jiang Z W, et al. Association Analysis of Cyber-Attack Attribution Based on Threat Intelligence[C]. *The 2017 2nd Joint International Information Technology*, *Mechanical and Electronic Engineering Conference*, 2017: 222-230.

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

[17] Nikolopoulos S D, Polenakis I. A Graph-Based Model for Malware Detection and Classification Using System-Call Groups[J]. *Journal of Computer Virology and Hacking Techniques*, 2017, 13(1): 29-46.

[18] Ding Y X, Xia X L, Chen S, et al. A Malware Detection Method Based on Family Behavior Graph[J]. *Computers & Security*, 2018, 73: 73-86.

[19] Cho I, Kim T, Shim Y, et al. Malware similarity analysis using API sequence alignments[J]. *Journal of Internet Services and Information Security*, 2014, 4(4): 103−114.

[20] Qiao Y C, Yun X C, Tuo Y P, et al. Fast Reused Code Tracing Method Based on Simhash and Inverted Index[J]. *Journal on Communications*, 2016, 37(11): 104-113.
(乔延臣, 云晓春, 庹宇鹏, 等. 基于 simhash 与倒排索引的复用代码快速溯源方法[J]. *通信学报*, 2016, 37(11): 104-113.)

[21] Tan Y, Liu J Y, Zhang L. Malware Familial Classification of Deep Auto-Encoder Based on Mixed Features[J]. *Netinfo Security*, 2020, 20(12): 72-82.
(谭杨, 刘嘉勇, 张磊. 基于混合特征的深度自编码器的恶意软件家族分类[J]. *信息网络安全*, 2020, 20(12): 72-82.)

[22] Alrabaee S, Wang L Y, Debbabi M. *BinGold*: Towards Robust Binary Analysis by Extracting the Semantics of Binary Code as Semantic Flow Graphs (SFGS)[J]. *Digital Investigation*, 2016, 18: S11-S22.

[23] Suarez-Tangil G, Tapiador J E, Peris-Lopez P, et al. Dendroid: A Text Mining Approach to Analyzing and Classifying Code Structures in Android Malware Families[J]. *Expert Systems With Applications*: *an International Journal*, 2014, 41(4): 1104-1117.

[24] Huang H, Youssef A M, Debbabi M. BinSequence: Fast, Accurate and Scalable Binary Code Reuse Detection[C]. *The 2017 ACM on*

Asia Conference on Computer and Communications Security, 2017: 155-166.

[25] Caliskan-Islam A, Harang R, Liu A, et al. De-Anonymizing Programmers via Code Stylometry[C]. *The 24th USENIX Conference on Security Symposium*, 2015: 255-270.

[26] Nataraj L, Karthikeyan S, Jacob G, et al. Malware Images: Visualization and Automatic Classification[C]. *The 8th International Symposium on Visualization for Cyber Security*, 2011: 1-7.

[27] Kothari J, Shevertalov M, Stehle E, et al. A Probabilistic Approach to Source Code Authorship Identification[C]. *The International Conference on Information Technology*, 2007: 243-248.

[28] Cavnar, William B. Using An N-Gram-Based Document Representation with a Vector Processing Retrieval Model[C]. *TREC*, 1994: 269-278.

[29] Burrows S, Uitdenbogerd A L, Turpin A. Application of Information Retrieval Techniques for Source Code Authorship Attribution[C]. *The 14th International Conference on Database Systems for Advanced Applications*, 2009: 699-713.

[30] Niculae D. General Unpacking: Overview and Techniques[D]. Sweden: Linnaeus University, 2015.

[31] Forrest S, Hofmeyr S A, Somayaji A, et al. A Sense of Self for Unix Processes[C]. *Proceedings 1996 IEEE Symposium on Security and Privacy*, 2002: 120-128.

[32] Bhatkar S, Chaturvedi A, Sekar R. Dataflow Anomaly Detection[C]. *2006 IEEE Symposium on Security and Privacy*, 2006: 15pp.-62.

[33] Christodorescu M, Jha S, Kruegel C. Mining Specifications of Malicious Behavior[C]. *The the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering*, 2007: 5-14.

[34] Quiring E, Maier A, Rieck K. Misleading Authorship Attribution of Source Code Using Adversarial Learning[C]. *The 28th USENIX Conference on Security Symposium*, 2019: 479-496.

[35] Zhang S, Zhou X, Yang F, et al. Traceback Mechanism Based on Neighbor Information in Wireless Sensor Networks[J]. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, 2013, 28(3):498-500.

[36] Shen X L, Shen J. Optimization Method of Tracing Distributed Denial of Service Attacks Based on Autonomous System and Dynamic Probabilistic Packet-Marking[J]. *Journal of Computer Applications*, 2015, 35(6): 1705-1709.
(沈学利, 申杰. 基于自治系统与动态概率包标记的 DDoS 攻击溯源优化方法[J]. *计算机应用*, 2015, 35(6): 1705-1709.)

[37] Zhao C. *Cognate mining and analysis of malicious code attacks based on network connection characteristics*[D]. Beijing: University of Chinese Academy of Sciences, 2016.
(赵灿. 基于网络连接特性的恶意代码攻击同源挖掘与分析[D]. 北京: 中国科学院大学, 2016.)

[38] Wu C X, Ma S L, Shi B, et al. Traceability of Network Attack Based on Electronic Fingerprint[J]. *Computer Engineering and Design*, 2020, 41(11): 3036-3041.
(吴朝雄, 马书磊, 石波, 等. 基于电子指纹的网络攻击溯源技术[J]. *计算机工程与设计*, 2020, 41(11): 3036-3041.)

[39] Wang J H, Chen Y L, Zhang Z Z, et al. Same Origin Attack Analysis Based on Features of Industrial Control System Function Code[J]. *Computer Engineering*, 2020, 46(7): 36-42.
(王建华, 陈永乐, 张壮壮, 等. 基于工控系统功能码特征的同源攻击分析[J]. *计算机工程*, 2020, 46(7): 36-42.)

[40] Wang Y D, Huang P, Jing T, et al. Research and Implementation on WebShell Comprehensive Detection and Traceability Technology Based on High-Speed Network[J]. *Netinfo Security*, 2021, 21(1): 65-71.
(王跃达, 黄潘, 荆涛, 等. 一种基于高速网络的 WebShell 综合检测溯源技术研究与实现[J]. *信息网络安全*, 2021, 21(1): 65-71.)

[41] Hutchins E, Cloppert M, Amin R. Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains[J]. *6th International Conference on Information Warfare and Security*, *ICIW 2011*, 2011: 113-125.

[42] Strom B E, Applebaum A, Miller D P, et al. Mitre att&ck: Design and philosophy[J]. *Mitre Product Mp*, 2018: 18-0944.

[43] APTnotes. https://github.com/aptnotes/data.

[44] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[EB/OL]. 2013: arXiv: 1301.3781. https://arxiv.org/abs/1301.3781

[45] Qaraei M, Abbaasi S, Ghiasi-Shirazi K. Randomized Non-Linear PCA Networks[J]. *Information Sciences*, 2021, 545: 241-253.

[46] Real R, Vargas J M. The Probabilistic Basis of Jaccard's Index of Similarity[J]. *Systematic Biology*, 1996, 45(3): 380-385.

[47] Huang A N. Similarity Measures for Text Document Clustering[J]. *New Zealand Computer Science Research Student Conference*, *NZCSRSC 2008 - Proceedings*, 2008: 49-56.

[48] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of Communities in Large Networks[J]. *Journal of Statistical Mechanics*: *Theory and Experiment*, 2008, 2008(10): P10008.

[49] Lambiotte R. Multi-Scale Modularity and Dynamics in Complex Networks[M]. Dynamics On and Of Complex Networks, Volume 2. New York, NY: Springer New York, 2013: 125-141.

[50] Verma R M, Zeng V, Faridi H. Data Quality for Security Challenges: Case Studies of Phishing, Malware and Intrusion Detection Datasets[C]. *The 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019: 2605-2607.

白波 于 2013 年在西安交通大学软件工程专业获得硕士学位。现在中国科学院大学网络空间安全专业攻读博士学位。现任北京网络数据研究所高级工程师。主要研究领域为网络攻防对抗、网络攻击溯源取证。Email: baibo_ucas@foxmail.com

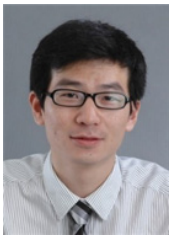
冯云 于 2021 年在中国科学院大学网络空间安全学院获得博士学位。现任中国科学院信息工程研究所工程师。主要研究领域为网络攻防技术、网络攻击追踪溯源。Email: fengyun@iie.ac.cn

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、中国科学院大学网络空间安全学院教授。主要研究领域为网络安全攻防对抗、态势感知、威胁情报、溯源取证等。Email: liubaoxu@iie.ac.cn

汪旭童 于 2020 年在中国矿业大学信息安全专业获得学士学位。现在中国科学院大学网络空间安全专业攻读硕士学位。研究领域为 Web 安全、网络溯源、机器学习。Email: wangxutong@iie.ac.cn

何松林 于 2020 年在成都信息工程大学信息安全专业获得工科学士学位。现在中国科学院大学电子信息专业攻读硕士学位。研究领域为网络空间安全、自动化安全研究。研究兴趣包括: 复杂电磁网络下的红蓝对抗、APT 攻击溯源与防御。Email: hesonglin@iie.ac.cn

姚敦宇 于 2015 年在福建师范大学网络工程专业获得学士学位。现在中国科学院大学电子信息专业攻读硕士学位。研究领域为 Web 安全、网络攻防。研究兴趣包括 Web 安全, 网络攻防。Email: yaodunyu@iie.ac.cn



刘奇旭 于 2011 年在中国科学院研究生院信息安全专业获得博士学位。现任中国科学院信息工程研究所研究员、中国科学院大学网络空间安全学院教授。主要研究领域为网络攻防技术、网络安全评测。Email: liuqixu@iie.ac.cn