面向组织溯源的威胁行为技术关联研究

陈柏翰[1]，王俊峰[1]，唐宾徽[2]，于忠坤[1]，葛文翰[1]，余　坚[1]
(1. 四川大学计算机学院，成都 610065；2. 四川大网络空间安全学院，成都 610065)

摘　要：在网络入侵日趋组织化的今天，如何有效地对威胁组织进行追踪溯源是网络安全防御中的重要内容. 威胁行为模式作为入侵受害者系统时的表现形式，于入侵者而言很难改变，是入侵者的一种高级特征. 若能有效提取组织的威胁行为模式，那么将大幅提高组织溯源的成功率和准确率. 为此，本文从组织行为模式的角度提出威胁行为技术关联算法. 该算法扩展了 Ward 连接凝聚层次聚类，可通过对组织所使用的入侵技术进行聚类学习以提取组织的威胁行为模式，并以 95%的置信度验证了威胁行为之间的技术关联性. 本文通过该算法生成的威胁行为技术关联模型，包含 97 类威胁行为技术关联簇，每一类簇可直观地看到不同组织所对应的威胁行为模式，可为组织溯源提供有力支撑.

关键词：网络入侵；溯源；威胁行为模式；技术关联；聚类
中图分类号：TP301.6　　**文献标识码**：A　　DOI：10.19907/j.0490-6756.2022.012006

Research on threat behavior technique association for organization traceability

CHEN Bo-Han[1], *WANG Jun-Feng*[1], *TANG Bin-Hui*[2], *YU Zhong-Kun*[1], *GE Wen-Han*[1], *YU Jian*[1]
(1. College of Computer Science, Sichuan University, Chengdu 610065, China;
2. School of Cyber Science and Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Nowadays, network attacks are becoming more and more organized. How to effectively trace the source of threat organizations is an important part of network security defense. As the manifestation of attacking the victim's system, threat behavior pattern is difficult to change, and it is an advanced feature of the attacker. If the threat behavior patterns can be extracted effectively, the success rate and accuracy of the organization traceability will be greatly improved. Therefore, this paper proposes the threat behavior Technique Association Algorithm from the perspective of organizational behavior pattern. The algorithm extends the Ward connection aggregation hierarchical clustering, which can extract the threat behavior patterns of the organization by clustering the attack techniques used by the organization, and verifies the technical correlation between the threat behaviors with 95% confidence. In this paper, the threat behavior Technique Association model generated by the algorithm includes 97 types of threat behavior Technique Association clusters. Each cluster can directly reflect the corresponding threat behavior patterns of different organizations, which can provide strong support for organization traceability.

Keywords: Network attack; Traceability; Threat behavior pattern; Technology association; Clustering

收稿日期：2021-06-15
基金项目：国家重点研发计划 (2018YFB0804503, 2019QY1400)；国家自然科学基金 (U20A20161, U1836103)；基础加强计划项目 (2019-JCJQ-ZD-113)
作者简介：陈柏翰(1996－)，贵州遵义人，硕士研究生，研究方向为网络空间安全. E-mail:cbhbob@qq.com
通讯作者：王俊峰. E-mail:wangjf@scu.edu.cn

1 引　言

奇安信威胁情报中心指出,在 2020 年的上半年,多个国外威胁组织利用新冠疫情相关热点事件对中国境内目标和机构实施高级可持续威胁(Advanced Persistent Threat, APT)活动[1]. 这些组织多次向我国各政要部门发起网络攻击,企图以各种方式渗透,以备日后开展破坏工作. 2020 年 12 月,美国网络安全公司 FireEye 遭遇黑客入侵[2],黑客成功窃取了 FireEye 的红队渗透测试工具,并搜索了与该公司某些政府客户有关的信息;同期,某黑客组织利用美国网络管理软件供应商 SolarWinds 的产品,并在其中植入木马后门,发动"SUNBURST"攻击[3],成功入侵美国财政部、商务部、国家电信和信息管理局(NTIA)等政府网络. 大量入侵事件表明,网络攻击组织化和入侵工具武器化已成为常态,如何有效抵御数量日益增多的网络攻击成为一个艰难挑战. 为了从根源上阻断网络攻击,人们往往需要追查入侵的源头. 对威胁组织的追踪溯源,已然成为网络安全防御建设中必不可少的内容.

然而随着入侵者攻击能力的不断提升,其躲避追踪的手段也日渐先进,如虚假 IP 地址、网络跳板、僵尸网络、隐蔽信道等方式在网络入侵事件中被大量使用,这些都给网络入侵的追踪溯源带来严峻的挑战. 传统的溯源技术往往只能获取局部的攻击信息,无法构建完整的攻击链条. 而在面向威胁组织的溯源中,攻击链条一旦中断,溯源工作则会无功而返. 如果无法确定攻击的根本源头,则难以防止攻击者的再次入侵.

研究表明,当入侵者成功进入受害者的系统时,往往表现出一致的行为模式[4]. 在面向组织溯源的过程中,威胁行为作为一种高级特征,于威胁组织而言很难改变,FireEye 公司安全专家 David Bianco 提出的"痛苦金字塔"模型[5]说明了这一点."痛苦金字塔是"对攻击方所使用的 IoC(Indicator of Compromise)进行分类和价值排序的层次结构,由下往上依次为哈希值、IP 地址、域名、网络特征或主机特征、攻击工具和 TTPs(Tactics, Techniques & Procedures). 于攻击方而言,由下往上改变对应 IoC 所付出的成本逐层递增,于防守方而言其利用价值也随之增加. 其中最顶层的 TTPs 是对攻击者攻击行为、战略战术层次的描述,它表示攻击者的行为模式,而调整 TTPs 对于攻击方而言所付出的时间和金钱成本最为昂贵. 因此基于 TTPs 的检测和响应会对攻击方造成更多的"痛苦".

文献[6]提出的使用基于 ATT&CK 的分析发现网络威胁方法指出,相比于入侵者通过修改 IoC 来躲避基于特征码的检测,结合行为分析的入侵检测能更有效抵御攻击,因为行为检测方法与入侵者对 IoC 进行的特定更改无关,而这也是推动 ATT&CK 不断发展的前提. 文献[7]提出的基于模糊哈希特征表示的恶意软件聚类方法,将恶意软件按一定的行为模式进行聚类,使得新型攻击更加容易被检测出来. 由此可见,作为入侵者的一种高级特征,行为模式的提取对威胁组织的追踪溯源意义重大.

基于以上原因,本文以 ATT&CK 知识库为基础,通过对组织威胁行为技术进行关联分析,从组织行为模式的角度提出威胁行为技术关联算法,以有效关联组织的威胁行为技术,并提取对应的 TTPs. 该算法结合了霍普金斯统计量和统计假设检验,以此扩展 Ward 连接的凝聚层次聚类. 本文通过该算法生成面向组织溯源的威胁行为技术关联模型,该模型具有以下几个特点:(1) 从攻击者的视角描述攻击过程,每一类覆盖了完整的 APT 攻击生命周期;(2) 弥补了 ATT&CK 中对于不同战术下技术之间关联性不强的问题,以 95% 的置信度水平证明了攻击技术之间的关联性;(3) 指导防御方通过采集和分析的方式来识别攻击行为,预测未检测到的攻击模式,并为组织溯源奠定理论基础.

2 相关工作

追踪溯源是网络安全领域的一种主动防御手段,是网络安全事故事后响应的重要组成部分. 按照攻击源头的不同,可将溯源划分为溯源攻击主机、控制主机、攻击者和威胁组织机构四个层次[8].

第一层对攻击主机的溯源技术在学术界得到了广泛研究. 如文献[9]中的概率包标记方法,主要思想为路由器以一定的概率给数据包做标记,当收到来自攻击方大量的数据包后,通过路径重构算法,重构其路径. 文献[10]是对概率包标记方法的改进,可以实现从少量数据包中重构攻击路径. Nur 等[11]提出的基于路由记录的 IP 追踪溯源方法利用 IP 协议的路由记录功能来实现概率数据包的标记,比其他技术需要的数据包更少. 文献[12]

提出了一种 IP 黑名单关联聚类算法，用于寻找恶意 IP 簇. 对攻击主机的溯源方法大多由包标记、日志存储等基本技术演变而来，其弱点也十分明显，攻击者若对 IP 地址等 IoC 进行伪装和隐藏，那么防御方则会无功而返.

第二层溯源的目标是定位控制主机. 相应的技术主要集中在对攻击者采取的不同隐藏技术进行溯源分析，分别是虚假 IP 的溯源、僵尸网络的溯源、匿名网络的溯源、跳板溯源和局域网的溯源[13]. 文献[14]提出了一种不可见的网络流水印技术 DROPWAT，用于在数据泄漏攻击中识别接收泄漏数据的暂存服务器. 文献[15]提出的一种多群粒子群优化算法，用于将数据包追溯到僵尸网络的攻击源. 同第一层溯源攻击主机，对控制主机的溯源也是在其入侵路径上进行追踪的过程，若要面向组织溯源，则仅从 IoC 层面出发是不充分的.

第三层和第四层是对攻击者和威胁组织机构的溯源，是在前两层溯源的基础上对网络空间和物理世界的信息数据加以分析，将网络空间中的事件与物理世界中的个人或组织联系起来的过程，这就要求追踪者必须找到网络威胁行为与攻击者或威胁组织之间的因果关系. 目前对第三和第四层次的溯源研究相对较少，但在网络威胁趋势逐渐国家化和军事化的今天，对威胁组织的溯源研究是十分必要的.

3　威胁行为技术关联算法

TTPs 描述了组织的行为模式，它通过对组织使用的技战术进行关联性分析得到. 本文设计了一种威胁行为技术关联算法，以实现对组织威胁行为技术关联性的分析和 TTPs 的提取. 详细流程如图 1 所示，它扩展了凝聚层次 Ward 连接聚类算法，根据数据的不同选择合适的距离度量，使用霍普金斯统计量进行聚类趋势评估，最后通过统计假设检验来确定截断值以完成簇的划分，并以此推断有意义的威胁行为技术关联.

3.1　距离度量

由于输入数据的性质不同，聚类相似度和距离度量的计算方式也存在差异. 常用的距离度量有欧式距离、曼哈顿距离、余弦相似度等，欧氏距离和曼哈顿距离计算的对象必须是连续变量，余弦相似度要求计算的样本在每一个维度上都要有取值.

通常威胁行为技术是多源、异构的非标准数据，其特征由离散的非对称二值变量组成，因此常用

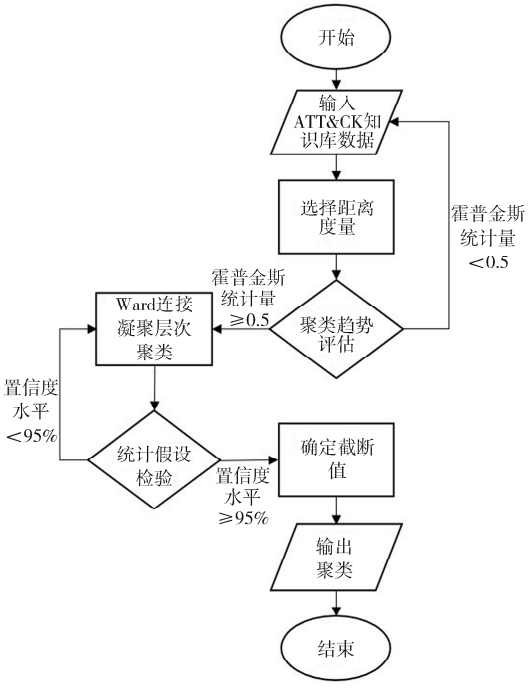

图 1　威胁行为技术关联算法流程

Fig. 1　Threat behavior technique association algorithm flow

的距离度量并不适用. 本文选择适用于非对称二值变量的 Phi 相似度来表示样本间的相似性. 任意两个技术 T_i 和 T_j 之间的 Phi 相似度定义如式(1).

$$Phi(T_i,T_j)=\frac{n_{ij}n_{i'j'}-n_{ij'}n_{i'j}}{\sqrt{n_i n_{i'} n_j n_{j'}}} \tag{1}$$

式中，n 代表数据集中威胁组织的数量；n_{ij} 表示同时使用过技术 T_i 和 T_j 的组织数量；$n_{i'j'}$ 表示都没使用过技术 T_i 和 T_j 的组织数量；$n_{ij'}$ 表示使用过技术 T_i 但没有使用过技术 T_j 的组织数量，同理 $n_{i'j}$ 表示使用过技术 T_j 但没有使用过技术 T_i 的组织数量. n_i 表示使用过技术 T_i 的组织数量，$n_{i'}$ 表示没有使用过技术 T_i 的组织数量；n_j 表示使用过技术 T_j 的组织数量，$n_{j'}$ 表示没有使用过技术 T_j 的组织数量，相关参数如表 1 所示.

表 1　相似度相关参数矩阵

Tab. 1　Similarity related parameter matrix

	$T_j=1$	$T_j=0$	总计
$T_i=1$	n_{ij}	$n_{ij'}$	n_i
$T_i=0$	$n_{i'j}$	$n_{i'j'}$	$n_{i'}$
总计	n_j	$n_{j'}$	n

Phi 距离 $d_{(P)}(T_i,T_j)$ 用于表示样本 T_i 和 T_j 之间的差异性，定义如式(2)所示.

$$d_{(P)}(T_i,T_j)=1-Phi(T_i,T_j) \tag{2}$$

式中，$Phi(T_i,T_j)$ 表示样本 T_i 和 T_j 之间的相似度.

3.2　聚类趋势评估

聚类趋势评估是评估数据集是否包含有意义的聚类,以及存在集群的数量.

对数据集进行聚类趋势评估是必要的,因为对于聚类而言,任何结构的数据集都能返回一个聚类的结果,即使其中的数据不相关. 本文采用霍普斯金统计量[16]来进行聚类趋势的评估,算法过程如算法 1 所示.

算法 1　聚类趋势评估

1) procedure Hopkins(D,n)
2) 从给定的数据集 D 中均匀的的采样 n 个点 $(T_1,\ldots,T_n)$.
3) for all $T_i \in D$ do:
4)　　$d_i \leftarrow \mathrm{dist}(T_i,T_j)$, 其中$T_j$是$T_i$的最近邻. 使用 Phi 距离度量计算 T_i和T_j之间的距离.
5)　end for
6)　以随机伯努利分布生成一个模拟数据集 D_R, 其中 n 个点$(R_1,\ldots,R_n)$的方差与给定数据集 D 相同.
7)　for all $R_i \in D_R$ do:
8)　　$\tilde{d}_i \leftarrow \mathrm{dist}(R_i,R_j)$,其中$R_j$是$R_i$的最近邻. 使用 Phi 距离度量计算 R_i和 R_j之间的距离.
9)　end for
10) 按照下列公式计算霍普金斯统计量:

$$H = \frac{\sum_{i=1}^{n} \tilde{d}_i}{\sum_{i=1}^{n} d_i + \sum_{i=1}^{n} \tilde{d}_i} \tag{3}$$

11) if H $>$ 0.5 then
12)　　存在有意义的聚类,执行聚类
13) else
14)　　不存在有意义的聚类,重新评估
15) end if
16) end procedure

3.3　Ward 连接凝聚层次聚类

与传统聚类算法不同,层次聚类使用指定的距离矩阵来创建基于树的数据表示. 虽然层次聚类算法已经存在很多种类型,但本文针对所使用的数据集设计了一种层次聚类算法. 该方法扩展了凝聚层次聚类和 Ward 连接方法.

凝聚聚类在寻找相似性最大的样本节点时具有出色的性能,因为它在融合决策的过程中会关注数据集的完整局部信息,因此,凝聚聚类会在样本最相似的地方创建更小的簇[17]. 在凝聚层次聚类的过程中,合并两个较大的簇需要用合适的连接方法来判断两个簇之间的距离. 本文使用 Ward 连接来推断威胁行为技术之间的关联性. Ward 连接使用离差平方和来度量两个簇之间的距离大小,当合并两个簇时,合并后的离差平方和会增大,选择使离差平方和的增量最小的两个簇合并,直到所有的簇合并为一个大簇为止,详细过程如算法 2 所示. 因此,Ward 连接可以使聚类后的簇更加紧凑,同时它对数据集中的噪声样本不太敏感[18].

算法 2　Ward 连接凝聚层次聚类

1) procedure Ward(D)
2)　for all $T_i,T_j \in D$ do:
3)　　$d_{i,j} \leftarrow \mathrm{dist}(T_i,T_j)$, 使用 Phi 距离度量计算$T_i$和$T_j$之间的距离.
4)　end for
5)　将T_i和T_j分别作为首行和首列,$d_{i,j}$为值生成距离矩阵 Dis_Matrix.
6)　for all $T_i,T_j \in$Dis_Matrix do:
7)　　按照下列公式计算离差平方和 ESS:

$$ESS = \sum_{i=1}^{n} T_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} T_i\right)^2 \tag{4}$$

8)　　将T_i和T_j合并,并按照式(4)计算合并后的离差平方和$\widetilde{ESS}$.
9)　　选择使$(\widetilde{ESS}-ESS)$的值最小的两个样本合并,直到最后合并成为一个簇.
10) end for
11) end procedure

3.4　统计假设检验

执行聚类之后,需要对结果进行检验. 最后的步骤分为两部分,第一是在树状图的树高上设置一个截断值,以创建最终的聚类集群,即簇的划分. 第二是验证技术之间的关联性.

本文采用统计假设检验的方法来对聚类结果进行检验. 用于统计假设检验的聚类树是由与原始数据集T_D具有相同的随机伯努利分布的数据生成的,用T_0表示. 如果T_D中的簇与T_0中的簇在指定的截止点上有很大的不同,那么学习产生的聚类树将产生具有统计意义的结果,而具有统计意义的层次聚类树说明了叶节点之间的关联性并不是随机产生的结果,这为推断技术关联提供了有效性. 算法 3 进一步解释了统计假设检验的具体过程.

算法 3　统计假设检验

1) procedure TEST(T_D)
2)　for possible $h \leftarrow 0.0$ to 10.0 do
3)　　$C \leftarrow$ 截断值为 h 时 T_D 聚类生成的簇的数量.
4)　　for $i \leftarrow 1$ to 1000 do
5)　　使用 Ward 连接凝聚层次聚类算法以随机伯努利分布生成一个层次树 T_0，使其方差与 T_D 的方差相同.
6)　　$C_i \leftarrow$ 截断值为 h 时 T_0 聚类生成的簇的数量.
7)　　end for
8)　　$p \leftarrow$ (count $C_i < C$) / 1000
9)　end for
10)　找到 p 中存储的最后一个值，使其 $\leqslant 0.05$，从而得出 95%置信水平下的统计显著性.
11) end procedure

4　算法仿真与实验结果

本文基于 ATT&CK 知识库数据，将 ATT&CK 技战术数据通过算法 1、算法 2、算法 3 依次执行完成聚类，并将生成的层次聚类结果与 ATT&CK 组织数据进行交集运算，以得到威胁行为技术关联模型.

4.1　数据源

本文使用的数据源为 MITRE 公司官方网站提供的开源数据集[19]. 该数据集由 MITRE ATT&CK 团队分析各类威胁情报所产生，是多源、异构的非标准数据集，具体数据如表 2 所示.

表 2　ATT&CK 数据集

Tab. 2　ATT&CK dataset

数据名称	数量/个
技术	178
子技术	352
威胁组织	110
恶意软件	518

4.2　聚类结果

本文首先通过聚类趋势评估算法得到数据集的霍普金斯统计量为 0.71，说明存在有意义的聚类簇；其次以 95%的置信度水平得到凝聚 Ward 连接的层次聚类树的最佳树高值为 2.6，即划分簇的截断值；最后，统计假设检验的结果也证明了技术关联的有效性.

最终的聚类结果如图 2 所示，纵轴表示截断值，横轴表示威胁行为技术的名称，共 97 个技术簇. 在截断值为 2.6 以下的部分，互相连接的技术归属到一个技术簇中，且具有较强的关联性. 部分细节展示如图 3 所示.

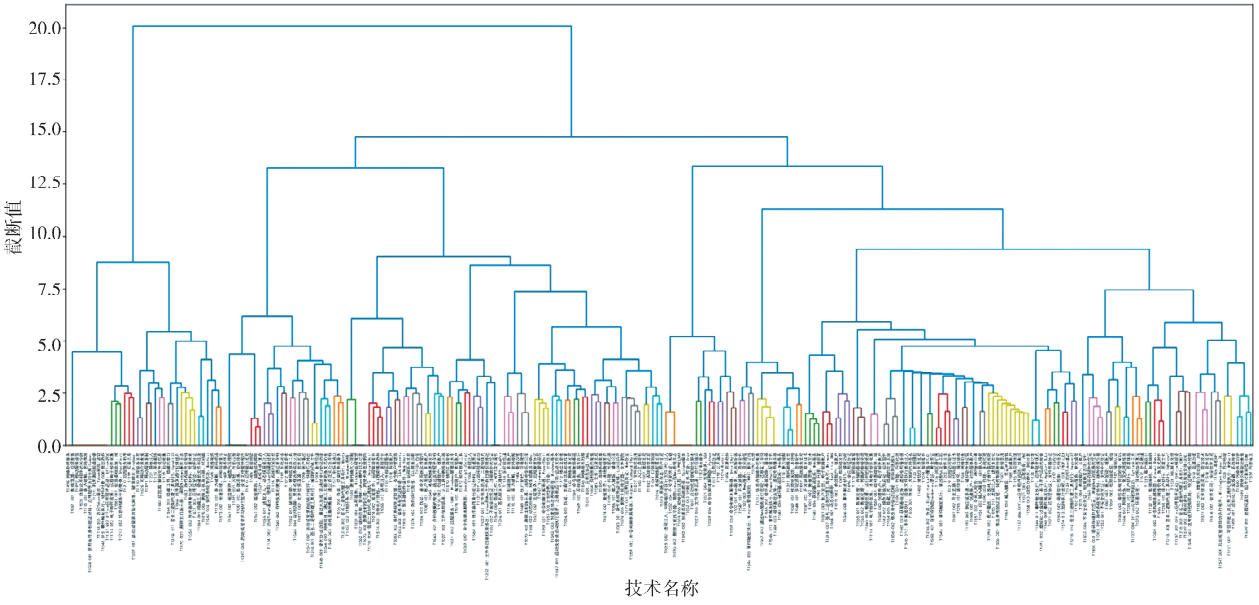

图 2　Ward 连接凝聚层次聚类结果

Fig. 2　Ward connection aggregation hierarchical clustering tree

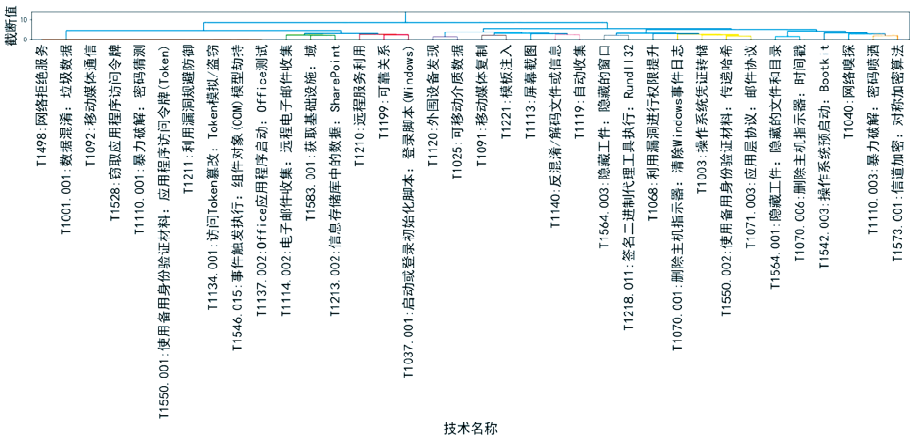

图 3　聚类部分细节展示

Fig. 3　Details of clustering

4.3　威胁行为技术关联模型

威胁行为技术关联算法对 ATT&CK 知识库进行仿真实验的结果是形成威胁行为技术关联模型，其重要特征是面向组织溯源，这要求模型中的威胁行为簇与威胁组织对应起来，以达到对组织行为特征的直观展示效果. 当溯源人员在对威胁技术进行分析时，可通过该模型找到包含该威胁技术的行为模式，并匹配到该行为模式所对应的威胁组织簇，最后通过必要的人工分析得出此次威胁事件的源头.

图 4　威胁行为技术关联模型可视化展示

Fig. 4　Visualization of threat behavior technology association model

本文将聚类结果与 ATT&CK 官方网站的威胁组织数据进行交叉运算，使聚类结果的每一个簇能够涵盖对应的威胁组织，以表示该类威胁行为主要出现在哪些威胁组织的攻击事件中. 本文使用 ECharts[20] 来进行模型的可视化展示，如图 4 所示，图 5 为部分细节展示.

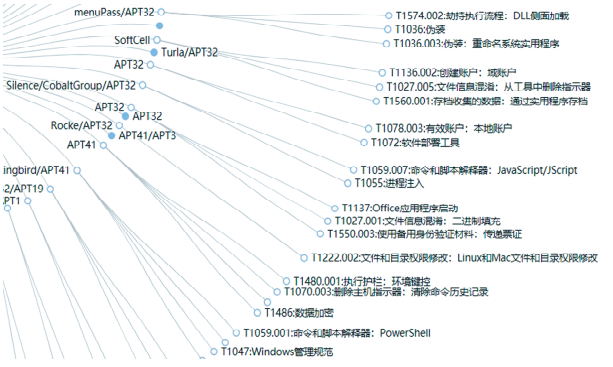

图 5　模型部分细节展示

Fig. 5　Details of the model

5　评　价

本节对威胁行为技术关联算法和提出的关联模型进行评价. 首先从统计学角度对算法进行有效性验证，其次对技术之间的关联性进行验证，再将关联模型与常见的威胁模型进行比较分析，最后结合专家经验综合说明本文算法和模型的有效性和创新性，具体评价指标如表 3 所示.

表 3　评价指标

Tab. 3　Verification index

验证指标	值
霍普金斯统计量	0.71
置信度水平	0.95
威胁情报覆盖率	0.86
标准覆盖率	0.88
专家经验	成立

5.1　技术关联性验证

在各大安全厂商所发布的一些威胁情报中，披露了攻击事件背后的威胁组织和使用的攻击手法.

本文收集了 1621 份威胁情报，这些威胁情报描述了一次攻击事件中攻击方所使用的手法，也就是本文所研究的威胁行为技术. 其中 1468 份报告中对威胁行为的描述是 ATT&CK 框架的战术和技术. 本文将这些包含 ATT&CK 技术的威胁情报整理出来，形成一份以技术为特征的验证集，并将它与本文所形成的威胁行为技术聚类结果进行交集运算，结果显示有 86%的报告与本文的威胁行为技术关联算法仿真结果吻合. 这也从实际应用上验证了技术之间的关联性是真实可信的.

5.2　与常见模型比较

为了说明算法仿真生成的威胁行为技术关联模型的创新性，共设计 8 个评价标准来与已有的威胁模型[21]进行横向对比，比较结果如表 4 所示.

表 4 中，评价标准分别为：C_1是否贯穿整个攻击生命周期，C_2不同攻击阶段是否含有攻击技术，C_3技术之间是否存在关联，C_4是否可预测未检测到的攻击技术，C_5技术之间的关联性是否直观，C_6是否支持定量分析，C_7是否能匹配威胁情报，C_8是否面向组织溯源.

表 4　模型标准对比

Tab. 4　Comparison of model standards

模型名称	提出年份	评价标准							
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
网络杀伤链模型	2011	✓	×	×	×	×	×	×	×
Mandiant 攻击生命周期模型	2013	✓	×	×	×	×	×	×	×
钻石模型	2013	×	×	×	×	×	×	×	×
MITRE ATT&CK 模型	2013	✓	✓	×	×	×	×	✓	×
Malone 杀伤链模型	2016	✓	×	×	×	×	×	×	×
Bryant 杀伤链模型	2017	✓	✓	×	×	×	×	×	×
Khan 杀伤链模型	2018	✓	✓	×	×	×	×	×	×
威胁行为技术关联模型	2021	✓	✓	✓	✓	✓	×	✓	✓

5.3　专家经验验证

在威胁行为技术关联算法仿真结果中，每一个类都涵盖了对应组织所使用的威胁行为技术. 本文通过专家经验知识，首先对威胁行为技术中存在的 TTPs 进行分析和验证，再通过两个具体的攻击事件综合说明仿真结果的有效性.

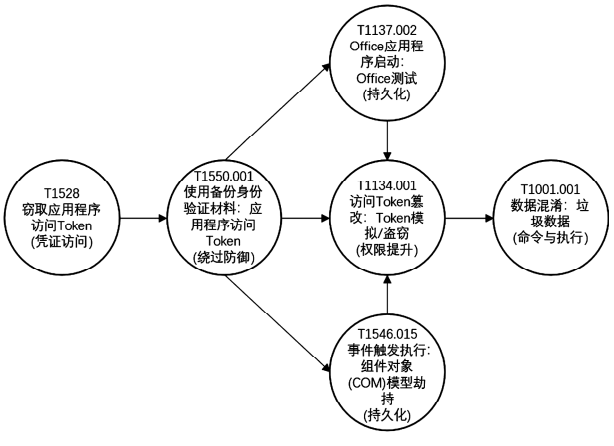

图 6　组织 APT 28 的一个 TTPs

Fig. 6　A TTPs of APT 28

组织 APT 28 的一个 TTPs 如图 6 所示，其在威胁行为技术关联模型对应的技术关联簇如图 7 所示.

图 7　在模型中对应的技术关联簇

Fig. 7　Corresponding technology association clusters in the model

APT 28 组织通过窃取应用程序访问 Token (T1528)技术来完成凭证访问，其次通过备份身份验证材料中的应用程序访问 Token(T1550. 001)来绕过防御，并使用两种方式来完成持久化的执行(T1137. 002，T1546. 015)，同时通过 Token 篡改来进行权限的提升(T1134. 001)，最后使用垃圾数据混淆来完成命令与执行操作，进而隐藏入侵痕迹，达到入侵目的. 可以看到，此 TTPs 是具有逻辑

解释性的,它描述了 APT 28 组织使用与 Token 相关的攻击技术来完成入侵目的.在模型中还存在类似的其他 TTPs,结合专家经验知识可以分析出威胁组织的入侵手段和目的.

2021 年 1 月 5 日奇安信公司发布的一则攻击模式分析报告中,披露了黑客组织"黄金狗"近期使用过的一些攻击手法[22],此组织通过误导、欺骗、伪装的方式使用户在不知情的情况下安装合法远控软件实现远程控制,其中伪装文档和诱饵文件成为此组织最常用的两个手法,并且通常会同时使用.在本文提出的威胁行为技术关联模型中,"网络钓鱼:鱼叉式附件"和"用户执行:恶意文件"两个技术的关联度非常高,它们通常会出现在同一起攻击事件中,这与奇安信所披露的"黄金狗"所使用的攻击手法近乎一致.

2019 年 4 月 5 日 FireEye 发布的一则威胁报告中,披露了金融 APT 组织 FIN 6 的一起入侵事件[23].分析人员发现 FIN 6 窃取数据的凭证,利用 Windows 的远程桌面协议(RDP)在环境中横向移动.在 RDP 连接到系统之后,FIN 6 使用两种不同的技术来建立立足点,第一种是使用 PowerShell 执行编码的命令以下载有效载荷,第二种是创建 Windows 服务来执行编码的 PowerShell 命令,最终下载有效恶意文件.在本文提出的威胁行为技术关联模型中,"远程服务:Windows 远程桌面协议""命令和脚本解释器:PowerShell"和"创建或修改系统进程:Windows 服务"这三项技术通常会同时出现,这与组织 FIN 6 的上述攻击事件吻合.

6　结　论

在面向组织溯源的过程中,行为模式揭示了威胁组织入侵行为之间的内在关联,是组织的重要特征.本文提出威胁行为技术关联算法,以 ATT&CK 知识库为基础对威胁组织的行为技术进行了关联性分析和提取,生成面向组织溯源的威胁行为技术关联模型,并建立了威胁组织的行为特征库和威胁技术库.实验结果表明,威胁行为技术关联算法能够有效提取威胁组织的行为模式,其生成的威胁行为技术关联模型能为网络安全防御人员提供有意义的溯源指导.

参考文献:

[1] 奇安信威胁情报中心. 全球高级持续性威胁(APT)2020 年中报[EB/OL]. [2021-04-28]. https://www.qianxin.com/threat/reportdetail?report_id=99.

[2] FireEye. Unauthorized access of FireEye red team tools [EB/OL]. [2021-04-28]. https://www.fireeye.com/blog/threat-research/2020/12/unauthorized-access-of-fireeye-red-team-tools.html.

[3] FireEye. Highly evasive attacker leverages SolarWinds supply chain to compromise multiple global victims with SUNBURST backdoor [EB/OL]. [2021-04-28]. https://www.fireeye.com/blog/threat-research/2020/12/evasive-attacker-leverages-solarwinds-supply-chain-compromises-with-sunburst-backdoor.html.

[4] 邬江兴. 网络空间拟态防御导论[M]. 北京: 科学出版社, 2017.

[5] Bianco D. The pyramid of pain [EB/OL]. [2021-04-28]. https://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html.

[6] Strom B E, Battaglia J A, Kemmerer M S, *et al*. Finding cyber threats with ATT&CK-based analytics[EB/OL]. [2021-04-28]. http://www.mitre.org/sites/default/files/publications/16-3713-finding-cyber-threats%20with%20att%26ck-based-analytics.pdf.

[7] 肖锦琦, 王俊峰. 基于模糊哈希特征表示的恶意软件聚类方法[J]. 四川大学学报: 自然科学版, 2018, 55: 469.

[8] 祝世雄. 网络攻击追踪溯源[M]. 北京: 国防工业出版社, 2015.

[9] Pfa B, Mt B, Mv A, *et al*. On packet marking and Markov modeling for IP traceback: a deep probabilistic and stochastic analysis[J]. Comput Netw, 2020, 182: 107464.

[10] Patel H, Jinwala D C. LPM: a lightweight authenticated packet marking approach for IP traceback [J]. Comput Netw, 2018, 140: 41.

[11] Nur A Y, Tozal M E. Record route IP traceback: Combating DoS attacks and the variants[J]. Comput Secur, 2017, 72: 13.

[12] 刘云, 肖添. IP 黑名单关联聚类算法对恶意簇检测的优化研究[J]. 四川大学学报: 自然科学版, 2021, 58: 013003.

[13] 姜建国, 王继志, 孔斌, 等. 网络攻击源追踪技术研究综述[J]. 信息安全学报, 2018, 3: 111.

[14] Iacovazzi A, Sarda S, Frassinelli D, *et al*. DROPWAT: an invisible network flow watermark for data exfiltration traceback[J]. IEEE T Inf Foren Sec, 2017, 13: 1139.

[15] Lin H C, Wang P, Lin W H, *et al*. A multiple-swarm particle swarm optimisation scheme for tracing packets back to the attack sources of botnet †[J]. Appl Sci, 2021, 11: 1139.

[16] Banerjee A, Dave R N. Validating clusters using the Hopkins statistic [C]// IEEE International Conference on Fuzzy Systems. Budapest, Hungary: IEEE, 2004, 1: 149.

[17] Manning C D, Raghavan P, Schütze H. Introduction to information retrieval [M]. Cambridge: Cambridge University Press, 2008.

[18] Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? [J]. J Classif, 2014, 31: 274.

[19] The MITRE Corporation. ATT&CK [EB/OL]. [2021-04-28]. https://attack.mitre.org/.

[20] Deqing L, Honghui M, Yi S, *et al*. ECharts: a declarative framework for rapid construction of web-based visualization [J]. VI, 2018, 2: 136.

[21] Ju A, Guo Y, Li T. MCKC: a modified cyber kill chain model for cognitive APTs analysis within enterprise multimedia network [J]. Multimed Tools Appl, 2020, 79: 29923.

[22] 奇安信病毒响应中心. 商业远程控制软件误导安装的黑产攻击模式分析[EB/OL]. [2021-04-28]. https://mp.weixin.qq.com/s/WpUKNXlQMzLAmDOO2uJiRg.

[23] FireEye. Pick-Six: intercepting a FIN6 intrusion, an actor recently tied to Ryuk and LockerGoga ransomware [EB/OL]. [2021-04-28]. https://www.fireeye.com/blog/threat-research/2019/04/pick-six-intercepting-a-fin6-intrusion.html.

引用本文格式：

中　文：陈柏翰，王俊峰，唐宾徽，等. 面向组织溯源的威胁行为技术关联研究[J]. 四川大学学报：自然科学版，2022，59：012006.

英　文：Chen B H, Wang J F, Tang B H, *et al*. Research on threat behavior technique association for organization traceability [J]. J Sichuan Univ: Nat Sci Ed, 2022, 59: 012006.