2020 | 巨人背后的专家

AISecOps智能安全运营技术白皮书

NSFOCUS

绿盟科技创新中心

绿盟科技威胁情报中心

天枢实验室
NSFOCUSTIANSHULAB

关于绿盟科技

绿盟科技集团股份有限公司（以下简称绿盟科技），成立于2000年4月，总部位于北京。公司于2014年1月29日起在深圳证券交易所创业板上市，证券代码：300369。绿盟科技在国内设有40多个分支机构，为政府、运营商、金融、能源、互联网以及教育、医疗等行业用户，提供全线网络安全产品、全方位安全解决方案和体系化安全运营服务。公司在美国硅谷、日本东京、英国伦敦、新加坡设立海外子公司，深入开展全球业务，打造全球网络安全行业的中国品牌。

版权声明

目录 CONTENTS

0 执行摘要

随着全球数字经济的蓬勃发展，网络安全与物联网、工业互联网、云计算、5G 等多种场景和技术进行融合，全面、深刻地改变了传统物理安全、生物安全、公共安全、国家安全等多层次的安全体系面貌。网络空间攻击面不断扩大，恶意攻击者规模化、组织化、军事化，攻击技术的自动化、智能化、武器化，在多种因素的作用下，在网络边界堆砌防护设备以期拒敌千里之外的思路已经失效。

面对日趋白热化、持续化的网络攻防对抗环境，安全运营（Security Operations，SecOps）面向人、技术、流程的集成与融合，提升安全防御资源的全局性、协同性，已成为安全能力落地，发挥防御体系有效性，对抗高级威胁的最直接、最关键环节。

可以预见，随着安全大数据的采集与智能分析技术的成熟，基于人工智能的安全运营技术方案（AISecOps）将大幅提升威胁检测、风险评估、自动化响应等关键运营环节的处理效率，大幅减少对专家经验的过度依赖，有效降低企业、组织乃至国家级关键信息基础设施、数据资产的整体安全风险。与此同时，智能安全运营技术能力的发展仍然在起步加速阶段，在体系架构、评估方法、数据融合、技术方向等多个层面，缺乏系统性的归纳与梳理。借此契机，绿盟科技推出《AISecOps 智能安全运营技术白皮书》，旨在对 AISecOps 智能安全运营技术的关键概念、成熟度、架构、技术等维度进行一个全面的总结与介绍，期望为读者带来全新的技术思考，促进 AISecOps 技术生态的构建，助力网络安全运营产业的技术升级。

本白皮书的主要观点如下：

- 对安全专家资源的需求与供给出现巨大剪刀差，安全运营智能化势在必行

传统专家驱动的安全运营，在数字时代，大规模安全运营数据接入的背景下难以为继，亟需提升安全运营的自动化水平。

- AISecOps 智能安全运营技术不是 AIOps（智能运维）、AISec（安全智能）、SecOps（安全运营）方案技术的简单叠加

AISecOps 在网络空间高度对抗环境下，面向安全运营风险管控的核心指标与关键环节，基于行为、环境、情报、知识等多维、多源数据，通过人 - 机智能融合，以全面提升安全运营能力的自动化水平。

- 当前 AISecOps 技术发展尚处于快速演进阶段，多项子技术的成熟度亟需升级

执行摘要

通过构建 AISecOps 技术框架，并形成技术成熟度矩阵，可以看到 AISecOps 多个阶段的关键技术能力远未成熟，技术研究任重而道远。

- 唯有可运营技术才能有效支持网络安全运营

数据智能驱动方法需要提升在安全语义适配、攻击意图理解、决策依据透明、深度人机互动等多方面的可运营属性，以提升机器智能与运营专家的数据、知识融合水平。

- 构建 AISecOps 技术“内功心法”图谱，对抗威胁的组织化、规模化、武器化

单点的、孤立的安全智能应用已经难以满足安全运营的系统化需求，通过构建细粒度的、场景化的、适当抽象的运营技术能力中台，支撑安全运营全生命周期智能化发展。

- 可信任安全智能是智能安全运营技术的未来

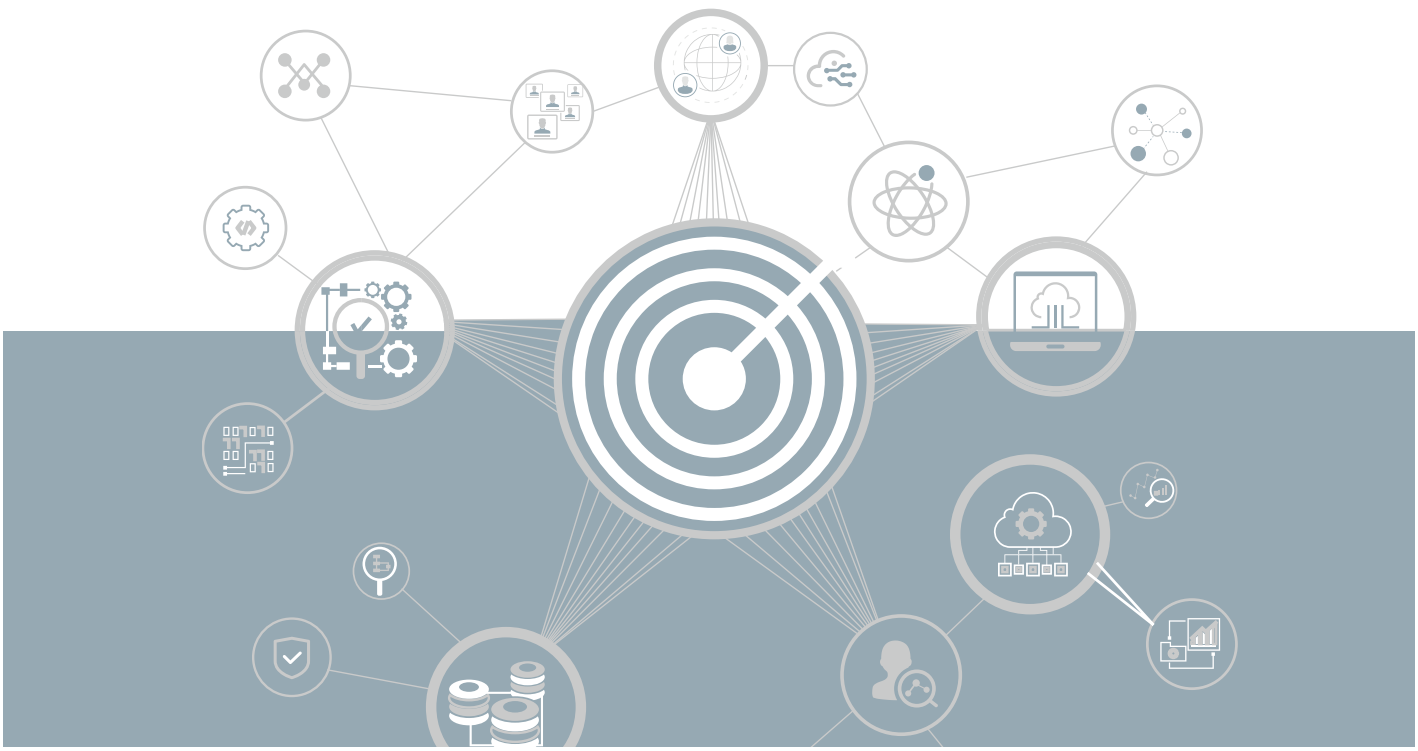
唯有高预测性能、透明可解释、安全鲁棒、合法合规的可信任安全智能，才能支撑网络安全运营中的关键决策输出，有效提升运营的自动化水平。

- 促进 AISecOps 技术生态建设，共建网络安全纵深防线

AISecOps 技术尚在起步发展阶段，需要技术生态的构建，促进相关标准的制定、数据与技术的共享、人才的培养，营造网络安全运营智能化大时代技术氛围。

1

安全运营发展背景与趋势



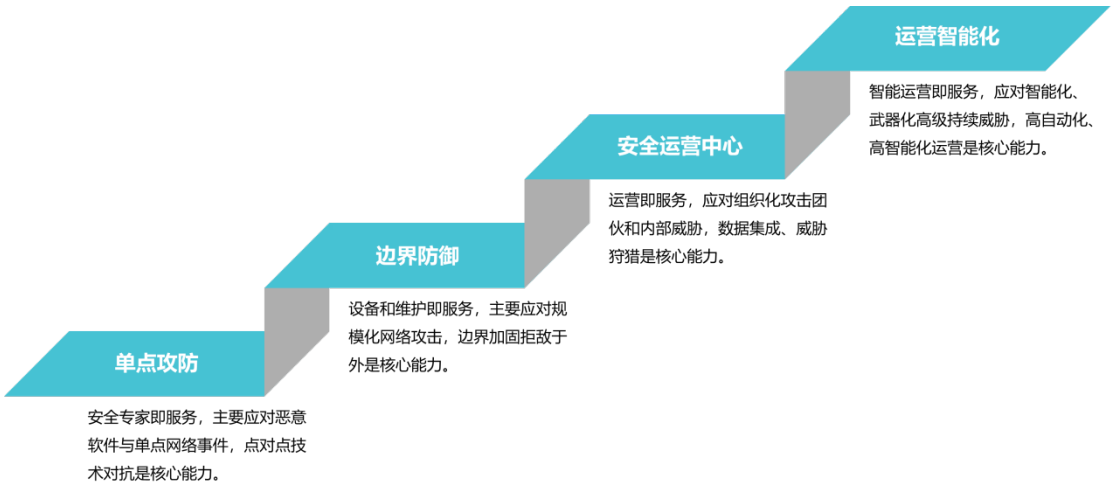


图 1　安全运营技术发展趋势

回顾安全产业的发展历程，在从计算机安全，到信息安全，到网络空间安全，再到数字安全，安全产业概念演进的背后，是网络信息化引领时代技术发展的核心趋势。然而，随着网络空间攻击面不断扩大，恶意攻击者持续规模化、组织化，攻击技术的自动化、智能化、武器化，多种因素的作用下，使得传统“高筑墙，广积粮”——在网络边界堆砌防护设备的被动防御思路逐渐失效。面对日趋白热化、持续化的网络攻防对抗环境，安全防御的思路开始不再局限在安全边界，逐渐形成更为成熟、更为完备的滑动标尺防护视角。边界防御的左移，是系统化的安全内生机制；右移是情报智能驱动的主动防御。零信任、威胁诱捕、威胁狩猎、安全开发、安全运营等支撑安全内生、主动防御的思路成为安全业界的共识。

安全视角的扩展，给网络空间安全防御方提供了完整的战略架构设计，然而，回到安全攸关的攻防战场，技术的演进更直接关系到战略实施的有效性。从技术演进的角度，攻防能力的较量已经逐渐演变为攻防参与者的军备竞赛。资源的投入决定了攻防双方的能力上限，然而防御之力有限，攻击之面无界，攻防信息的严重不对称性，决定了安全防守方处于被动的局势之中。如何在有限的信息、资源下，充分覆盖安全滑动标尺的能力范畴，有效降低企业、组织乃至国家的系统性安全风险，打赢网络空间的渗透战、攻坚战，已成为全面数字时代网络安全的最关键目标。

安全风险的感知、监控、预警、处置、评估等需要系统性、持续性的运营机制来保证，正是这个驱动力，使得安全运营成为整个安全防御生命周期至关重要的组成和技术发展的热点。安全运营能力的成熟度，也已成为决定整个安全基础架构、安全防御机制、安全防护设备、安全研究技术能否有效发挥作用的最

关键因素之一。近年来，安全运营中心、可管理的安全运营服务、威胁情报运营等运营技术设施驱动了新一轮的安全产业发展。于此同时，安全运营技术，包括威胁识别与检测、事件预警、事件溯源、自动化响应等等，也是安全研究领域的热点，围绕安全运营的流程构建、技术实现、人为因素等多方面的研讨活动层出不穷。正如近现代战争的成败很大程度上决定于现有战略资源的调度能力，安全运营能力，实际上是技术、人、流程等多环节安全能力的综合体现，是协调、调度、整合安全资源实现网络空间防御的关键环节。**无论是安全左移追求安全机制内生，还是安全右移促进主动安全防御，安全运营愈发成为安全能力内外兼修的必由之路。**

安全运营（Security Operations, SecOps）的关键在于，通过流程覆盖、技术保障及服务化，为企业、组织等实体提供脆弱性识别与管理、威胁事件检测与响应等安全能力，以充分管控安全风险[1]。**安全运营中的概念核心就是管理风险，而风险的度量是动态的、持续的、相对的。**在不同的网络环境下、在不同的攻防场景下、在不同的资源配比下、在不同的风险偏好认知下，风险具有不同的表现形式。风险难以消灭，在有限的资源投入下，企业或组织能够对安全风险可识、可管、可控，则表明其具有弹性的（Resilience）、健壮的（Robustness）安全能力。

正是由于安全运营风险驱动的特性，对风险的认知的演进，决定了安全运营认知的方向。整体来看，安全运营技术和产业经历了单点攻防、边界防御、安全运营中心的发展历程，并最终向运营智能化的方向持续演进。

- **单点攻防** 伴随着互联网时代的到来，针对个人电脑的恶意软件率先爆发。网络世界的威胁趋势逐渐呈现在大众面前。此时恶意软件正是最大的安全风险，大量的攻防专家开始投入到反病毒软件的研制当中。安全运营的概念还未成型，专家即服务是典型的安全能力交付方式。
- **边界防御** 利益驱动之下，攻击与威胁逐渐组织化、产业化；于此同时，大规模互联网服务与 IT 系统软件的迅速演进，软件漏洞引发的安全脆弱性问题浮出水面。为此，抗 DDoS 攻击、入侵检测系统、远程漏洞扫描系统应运而生，快速构建起网络防御边界。并随着攻防研究的深入，威胁场景的快速迭代，此时的安全运营从萌芽到成长，渗透测试、风险评估团队的配套逐渐成型，设备和维护即服务成为主流。
- **安全运营中心** 高级持续性威胁（Advanced Persistent Threat，APT）和相关事件的出现，给边界化防御的思路带来巨大的冲击；于此同时，多层次的安全政策、规范形成合规性要求。多个因素的汇集形成整个网络空间认识的颠覆性变化。常态化、协同化、纵深化和智能化的防御思

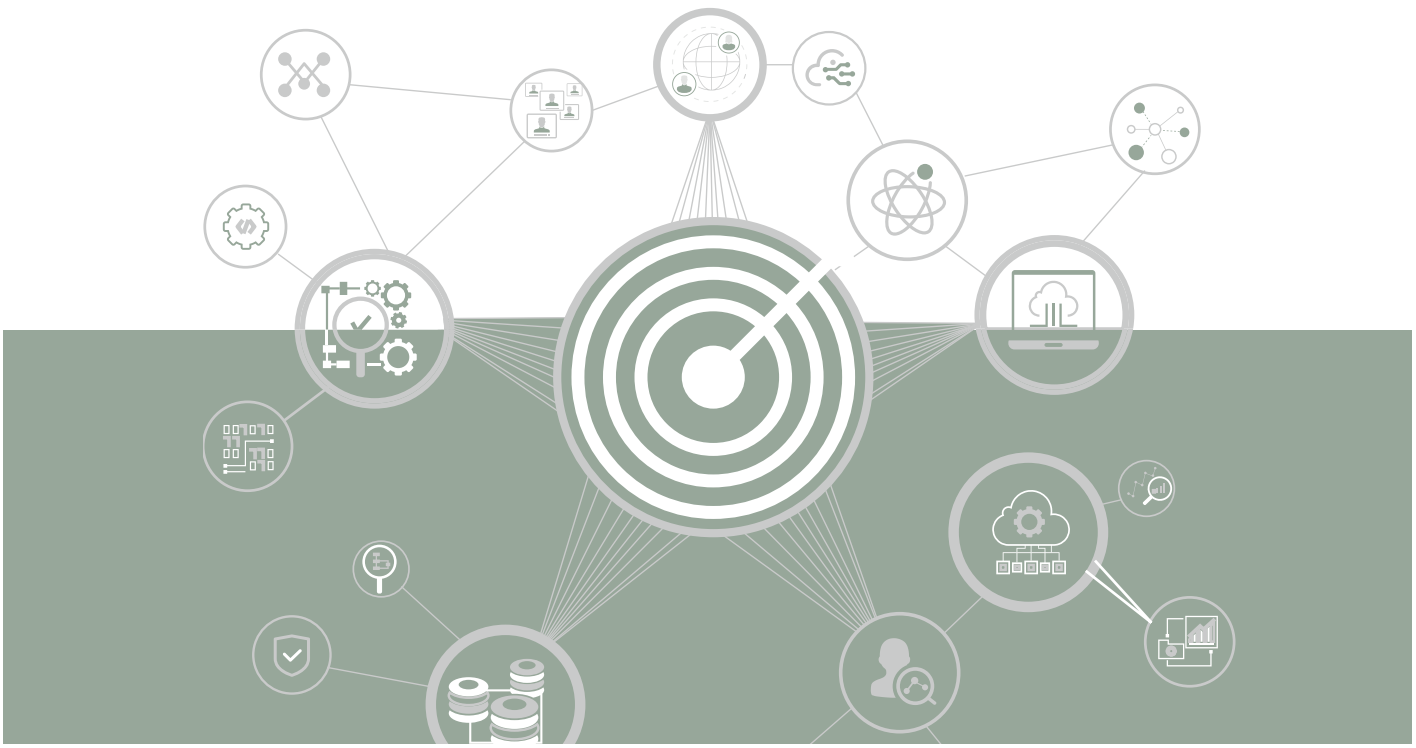
路成为业界共识。此时，安全运营理念和架构逐渐成型，安全运营中心（Security Operations Center，SOC）遍地开花，以中心化的方式管理威胁、脆弱性、资产等风险相关的流程和数据，并辅以行为分析、蜜网诱捕、威胁狩猎、情报融合等高级安全技术，提升安全运营的效率。运营即服务，正成为当前网络空间防护的关键趋势。持续自适应风险与信任评估（Continuous Adaptive Risk and Trust Assessment，CARTA）架构与理念，也正是在这个背景下得以普及。

- **运营智能化** 安全运营团队，是支撑安全运营中心化运作的核心。安全运营的萌芽、发展与成熟，映射出的是背后人与人对抗的认知与技术升级。然而，随着网络空间对抗关联流程链路的增长、数据规模爆炸、技术复杂度提升，人力资源与风险管控的目标要求之间，逐渐形成巨大的需求剪刀差。这种数字化时代的关键特征，倒逼网络安全运营突破依赖安全专家的传统“人工”阶段。提升安全运营技术与流程的自动化、智能化水平，已成为网络安全风险治理与防控的必备条件。智能赋能运营，是数字化时代运营即服务的基础保障。

安全运营智能化趋势已成为必然。流量分析、行为分析、样本分析、威胁关联、自动化响应等技术越来越多地采用了机器学习算法、图算法、强化学习算法，尽管如此，现阶段安全智能的发展水平，仍难以满足安全运营对威胁发现实时性与准确性、事件自动化溯源、风险决策自动化等多方面的要求。距可用、成熟的智能安全运营服务，我们还有很长的路要走。

2

智能安全运营的挑战



数字时代的背景下，数据和智能驱动的安全对抗，技术平台的自动化、智能化水平，愈发成为网络空间中攻防双方角力的重点。回归到攻防的战场上，我们希望能够得到的是一个能处理海量异构多源数据，快速检测、溯源和预测威胁事件，辅助安全团队进行分析、推理、处置的自动化安全运营平台。

本质上，安全运营中大规模数据分析的困难来自于攻守的不平衡性。可持续的安全运营的目标是在合理的投入产出比下，持续的监控并降低企业和组织的系统安全风险。安全运营的目标不仅仅是在态势感知大屏上看到威胁趋势，而是真正要发现并处置真实威胁，例如进行针对高隐匿性、低频的高级威胁的威胁狩猎。

安全运营智能化，可借助人工智能和自动化编排等技术，有效提升安全运营能力的自动化水平，降低威胁分析与响应的周期，减少对人力投入与专家经验的依赖，简化安全运营流程。

但在真实的网络空间中，敌暗而我明，智能安全运营需要大规模地采集多维度的数据进行分析，但处理海量数据给安全运营团队带来了前所未有的挑战，如依赖爆炸、告警疲劳、大海捞针（威胁）等难题，都可能是整个运营团队的梦魇。除此之外，技术瓶颈，专业人才匮乏，流程低可操作性等问题，都将降低安全运营的有效性。

以下，从网络安全运营关键实用性的角度，总结安全运营中大数据带来的关键技术挑战。

2.1 数据接入：数据膨胀与系统瓶颈

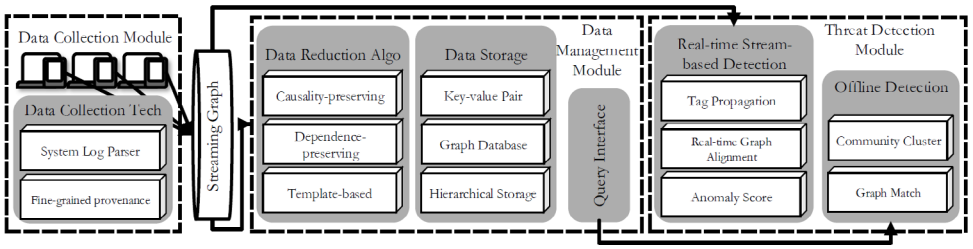


图 2　溯源数据分析系统的一般技术框架 [2]

如前所述，一方面高级威胁低频且具有隐匿性，另一方面企业和组织需要持续进行风险管控。因此，为全面的评估系统风险，所需采集的数据种类多、数据规模异常庞大。以终端侧日志分析为例，图 2 展示了一个典型终端威胁检测处理系统的架构，涉及从数据采集、数据管理、威胁检测等多个环节。如果没有有效的预处理环节，单台用户主机的日常流量、终端行为日志量至少每天可达数百兆字节，更不用

说提供服务资源等功能性节点。不止是数据吞吐量大，为了满足合规需求，支持事件溯源、关联等威胁分析任务，所采集的数据往往需要长达数百天的持久化留存。这些数据的采集、传输、存储等给计算、网络、数据库等各个系统环节带来巨大的压力。其衍生后果就是，许多采集能力被禁用，大量数据在预设的价值判断策略下被提前丢弃，这可能导致威胁线索和证据链的时效。数据爆炸所产生的这些现实问题成为 XDR（eXtended Detection and Response）等技术方案落地的关键阻碍。

2.2 数据融合：多源异构与本体建模

大规模数据分析需要终端侧、网络侧、沙箱侧、蜜罐侧的日志告警，以及威胁情报、知识库、IT 资产、扫描的漏洞、HR 信息等多源异构的数据，且依赖高层次数据模式的融合，一个典型的本体设计防范如图 3 所示 [3]。现阶段欠设计、低耦合、低交互的数据集成造成了数据爆炸，难以建立高质量的融合数据基础。多源多维数据规范化、本体化、体系化，始终是智能分析技术的基石。当前，各类不同厂商的网络安全设备执行不同的数据命名、标注策略，亟需在统一的语义下实现数据接口的统一规范化，以实现低成本的数据集成与交互。同时，多源异构数据中包含大量关联实体、重复实体，为实现这些数据实体的一致性关联分析，需要以全局的视角，将数据抽象本体化，设计体系化数据模型。例如，以图模型整体建模实体节点及实体间的交互行为，能够自然利用网络数据的关联属性，并进一步应用多种图分析策略与方法。

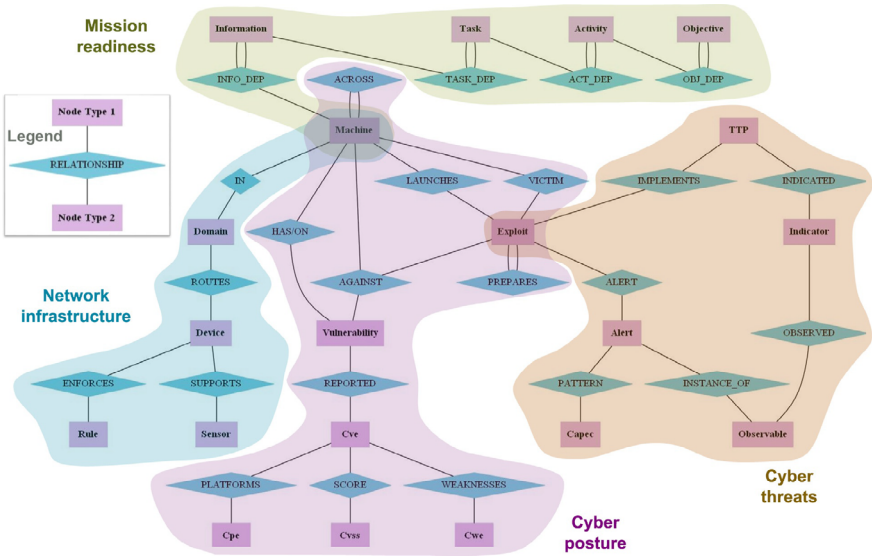


图 3　CyGraph 本体设计

2.3 线索发现：召回模型与高误报率

传统基于静态规则的安全策略，难以快速应对高度产业化、武器化的攻击技战术快速演进。行为分析、意图分析、用户画像等技术，通过多维度的数据挖掘建立用户行为基线、提取行为模式，能够有效弥补传统静态方法的不足。然而数据驱动的威胁线索识别，仍然逃不出高误报率的魔咒。以 ATT&CK（Adversarial Tactics, Techniques, and Common Knowledge）驱动的行为分析为例，该矩阵中的大部分攻击技术抽象都是召回策略驱动的。如下图所示，是 MITRE 所跟踪观测的 93 个 APT 组织利用次数最多的十种技术 [4]（该技术划分命名基于改版之前的 MITRE 矩阵，尚未包含子技术的概念）。其中能够直接对应到攻击行为的技术描述，只有鱼叉式网络钓鱼（Spearphishing Attachment），凭证窃取（Credential Dumping）和文件混淆（Obfuscated Files）这三类，其他七类技术划分单独来看，都是正常网络行为与操作。ATT&CK 的关键目标在于覆盖和召回，而从安全运营的视角来看，在事件规模膨胀的现状下，误报率是一个非常关键的有效性衡量指标。一项针对赛门铁克终端告警的分析表明，由 34 台机器触发的 58096 条告警中，与检测目标 APT29 行为相关真实告警只有 1104 条，告警的精度只有 1.9%。海量告警场景下高误报告警带来的误报疲劳，会最终降低整个安全运营团队的运转效率 [4]。

误报不止是召回模型的模型设计本身引入的，在机器学习的统计建模过程中，样本空间的不对称性，训练数据与实测数据的分布偏差等多方面的因素，会进一步导致模型预测性能在实际运行中的大幅衰减，同样会产生大量误报。

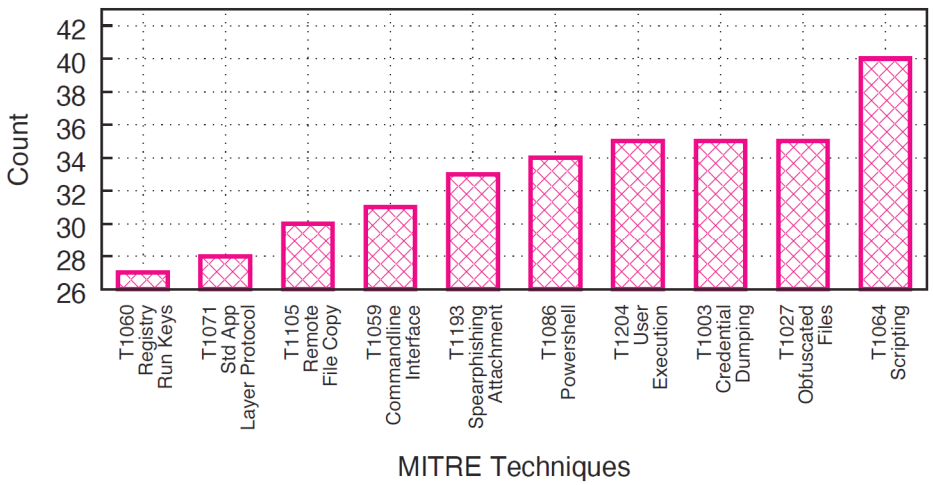


图 4　MITRE APT 关联的常见技术统计

2.4 事件推理：语义模糊与依赖爆炸

安全事件日志是网络实体的高层次目标与具体行动的数据层次映射，具有潜在的行为层次语义化、结构化关联与逻辑依赖关系。仍以 ATT&CK 驱动的威胁检测为例，通过阶段划分，给具体技术的归类赋予了一定的语义关联，为安全团队讲故事提供了线索串联的范本。然而，从数据挖掘和关联的角度，有两个重要的问题需要考虑。第一个问题是一词多义，是指一个技术可能横跨多个战术实现，并以不同的粒度出现在一定的威胁上下文中[5]。例如 T1053 定时任务（Scheduled Task/Job），包含在执行（Execution）、持久化（Persistence）和提权（Privilege Escalation）三个战术目标中。ATT&CK 将 T1053 技术划定为一种统一的技术，并未针对具体战术进行细粒度的描述。这本质上是由 ATT&CK 的技术抽象层次决定的，然而这给数据分析任务带来新的挑战——需要解决充分理解技术触发的上下文，并赋予该技术明确的战术语义。

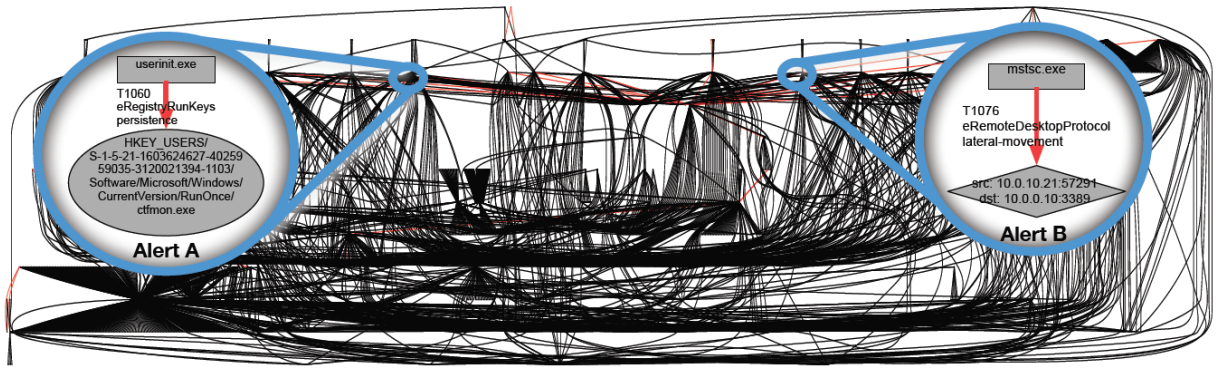


图 5　APT 29 攻击溯源数据图[4]

第二个问题是依赖爆炸。这包含两个层次，第一个层次是 ATT&CK 的战术模型不是因果模型，也不具有统计意义。我们可以从 MITRE 提供的 APT 实例中看到具体的技战术执行数据流。然而，在实际检测、溯源分析中，技战术的跳转是矩阵中的多战术之间、单战术之内的多种技术方案的排列组合问题，在任何特定场景和实际环境中的高级威胁行为序列是独特的，规律性难以捕获。第二个层次是在细粒度的溯源数据层面（Provenance），现阶段的数据采集在一定的资源限制下，难以精细刻画信息传递流。像文件操作、网络输入、进程创建等，存在一对多、多对多的路径依赖问题。由于该层次数据的细粒度特性，依赖爆炸直接加剧了数据存储、检测、溯源等各个环节的技术难度。

2.5 人机协同：黑盒模型与低质交互

当前阶段，网络安全运营关键环节的决策主体仍然是人。安全运营平台需要建立与运营人员的沟通机制，以有效实现人机智能协同。如图 6 所示，基于深度学习等复杂不可解释的黑盒模型，以及低交互甚至无交互的人机交互流程设计，是人机协同机制构建的重要阻碍。在数据驱动的应用场景下，人工智能系统需要以足够透明、可解释的方式输出其判断逻辑和决策过程。不可信任的人工智能，显然不能够胜任任何对系统安全和人身安全攸关的关键性场景，这将大大降低其可用性和适用范围[6]。在网络安全运营的场景下，黑盒人工智能模型，所提供的识别、检测结果，甚至是推荐策略，如果不能够提供人能理解的、可供审计的判断解释依据，将无法被集成到自动化的运营流程当中去。

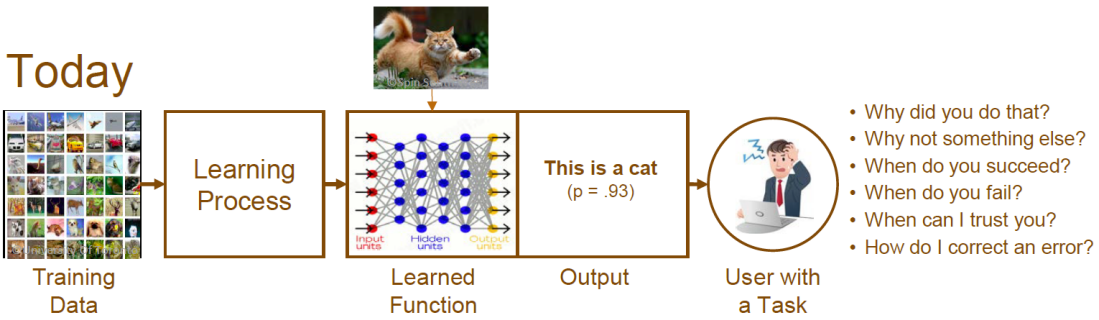


图 6 不可解释的黑盒模型[7]

除此之外，需要有效的反馈接口、理解引擎，将人类的反馈信息融合到实时调整的模型参数和结构当中去。当前安全运营中心所依赖的 SIEM、SOAR 等平台，绝大部分都是被动的记忆模式——机械的记录输入的规则和历史剧本。这种低泛化或无泛化能力的机制无法有效实现真正的人机智能融合。

2.6 智能引擎：攻击失效与数据风险

人工智能自身的安全性问题，同样是安全运营智能化数据分析实践应用中不可回避的挑战之一。当前针对人工智能模型与算法的攻击技术频出，通过对抗样本等手段可诱发错误的机器判断。结合安全语义语法规则，对抗样本、对抗载荷能够绕过防护设备的检测与分析，甚至导致模型对实时基线的误判，造成对正常业务的误杀。保证智能安全运营系统组件的安全鲁棒性，需要安全及数据分析团队在模型、算法构建之初充分考虑。

此外，如图 7 所示，智能化引擎的训练、识别过程可能涉及企业安全运营中的个人隐私数据与企业

敏感数据，攻击者可通过参数推断、模型窃取等技术手段实现数据盗取。因此数据安全性也已成模型落地过程中的关键考量因素，以降低智能化技术引入的伴生数据风险。

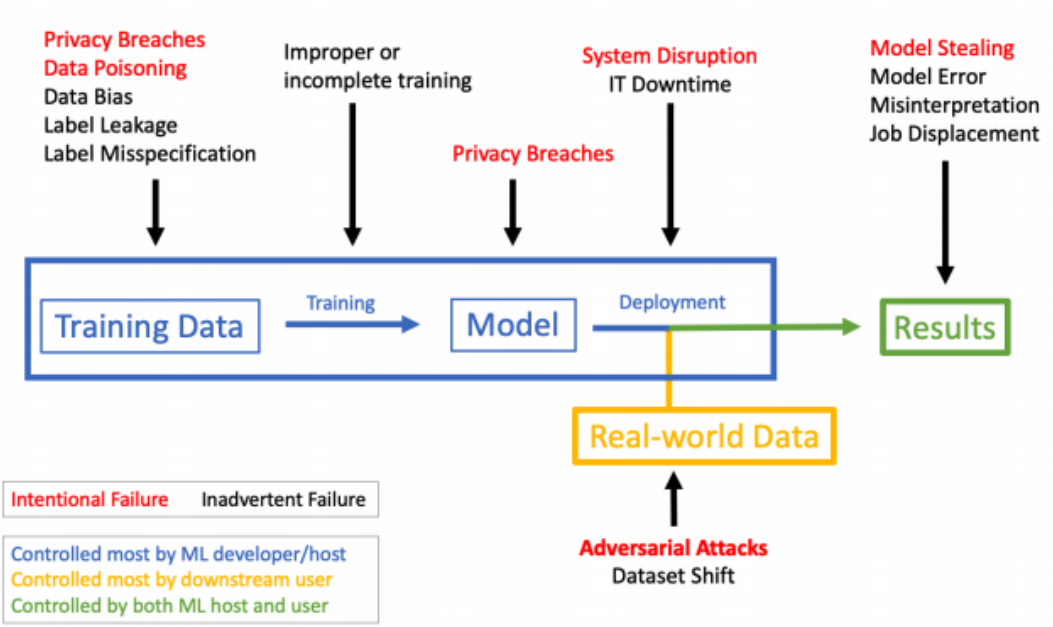
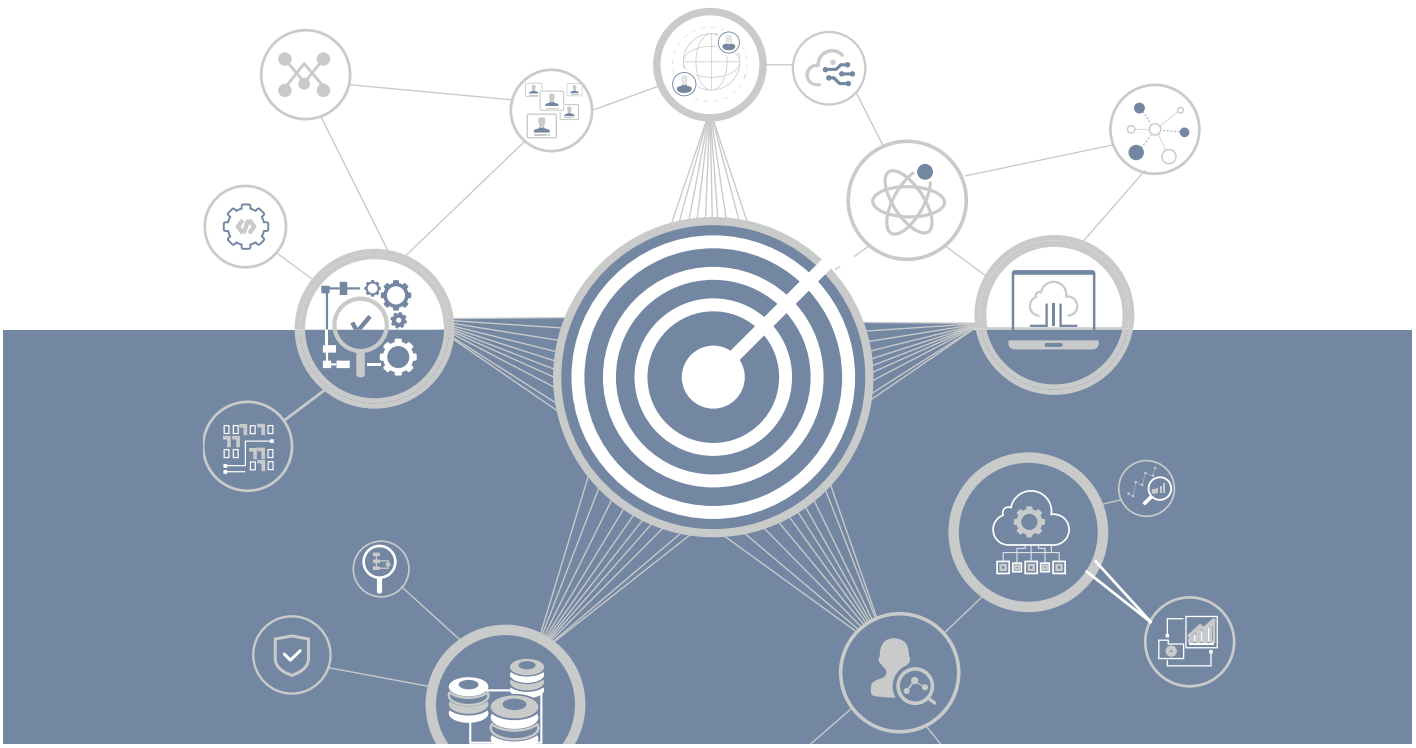


图 7　人工智能模型的主要攻击面 [8]

3

AISecOps智能安全运营技术体系



AISecOps 技术发展尚处于初级阶段，本章将全面梳理、总结、提出 AISecOps 的技术内涵、指标体系、数据分类、技术框架、成熟度矩阵等关键概念和理论，以明确 AISecOps 技术的发展方向与技术路线。

3.1 AISecOps 核心内涵

从基本的词语组合来看，AISecOps 由 AISec，SecOps，AIOps 三大核心技术组成。

人工智能安全（AISec）的技术融合给行业带来了新的期盼。无论是 AI 自身安全还是基于 AI 的安全应用，都已成为学术界和工业界的热点话题。AI 技术在诸多单点安全技术和指定场景中，如恶意软件分类、恶意流量识别、入侵检测等，已取得不错的应用效果。

IT 智能运维（AIOps）亦是整个互联网、智能计算领域的研究热点 [9]。该技术方向重点关注复杂 IT 系统环境的异常检测、根因定位、告警分诊等关键技术。不过，IT 运维不同于安全运营，缺乏对网络威胁、脆弱性、资产等核心风险要素的系统化建模，相关技术经验难以直接复用到安全运营场景中。

最后，安全运营（SecOps）作为应用场景与目标，主要由流程、人和技术三个核心要素构成。本文更关注技术要素在智能安全运营时代下的发展趋势。传统安全运营的技术能力主要由安全专家提供，例如告警分类分级、威胁狩猎、样本分析、威胁溯源等等。然而，基于安全专家的运营能力与快速膨胀的防护需求之间，已逐渐形成巨大的剪刀差，安全人才的缺口与瓶颈日趋严峻。因此，探索智能安全运营技术方案，已是迫在眉睫。

图 8　AISecOps 核心技术能力拆解

AISecOps，智能安全运营技术并不是简单地 AISec、SecOps 和 AIOps 技术的加和。人工智能赋能的有效性，一方面取决于人工智能技术自身的发展水平，另一方面，更决定于人工智能技术与相关应用场景在核心目标、体系架构、功能需求、数据模型的多方面的融合程度。智能化是手段，而不是目标。为此，本文归纳了智能安全运营的核心内涵，以明确技术实现与发展的范畴：

“AISecOps 技术是以安全运营目标为导向，以人、流程、技术与数据的融合为基础，面向预防、检测、响应、预测、恢复等网络安全风险管控、攻防对抗的关键环节，构建数据驱动的、具有高自动化水平的可信任安全智能技术栈，实现安全智能范畴下的感知、认知、决策、行动能力，辅助甚至代替人在动态环境下完成各类安全运营服务。”

不同于 AISec 实践中智能化技术与安全领域的单点结合，智能安全运营是在核心运营指标的导向下，系统、深入的多维融合智能化技术方案，以适应安全运营不同阶段、不同任务场景的应用需求，这对传统人工智能技术的鲁棒性、可信性、安全性提出了全新的要求。

3.2 AISecOps 指标体系

以上内涵中，安全运营目标是引导技术能力发展方向的关键。因此，本文面向安全运营的关键需求，自顶向下总结了 AISecOps 技术的指标层次，分别包括愿景目标、运营指标和技术指标。其中，技术指标又可分为数据指标和分析指标。

在此，我们只粗略地划分了指标体系的层次，列举不同层次中的一些关键指标。值得注意的是，在层次化指标设计的基础上，需要精细化设计指标的依赖性关系与数值化度量，以促进指标体系的有效运转。

网络安全运营能力的提供是目标导向的。从企业、组织、国家的愿景目标出发，进而构建安全运营任务级别的运营指标，进一步指导构建数据与分析层面的技术指标，最终形成图 9 所示的层次化指标体系，以评估技术实现的有效性。

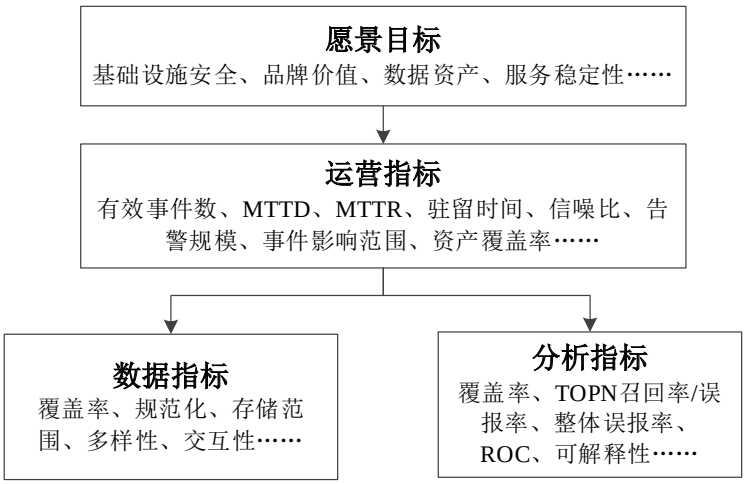


图 9 AISecOps 指标体系

愿景目标指企业、组织、国家等主体层面的核心安全、业务、商业目标。例如，维护 IT 基础设施的稳定运行，保护核心数据资产，维护品牌价值的安全性等。这些愿景目标与主体的发展目标密不可分。

运营指标以愿景目标为基础，针对网络安全相关的业务能力制定安全运营核心指标，以评估安全运营能力水平。在运营指标的导向下，需要有针对性地对数据融合水平和分析技术水平进行评估，以促进技术能力的迭代。

在数据层面，需要考虑包括覆盖率、规范化、存储时效、多样性、交互性等指标；在分析层面，不仅要考虑传统机器学习等技术的评估指标，包括预测精确性、召回率、ROC 等，还重点考察场景覆盖率、TOPN 召回率 / 误报率、整体 / 单点误报率及模型可解释性等面向可运营、易运营的分析指标，合理促进技术与人、流程的深度融合。

3.3 AISecOps 数据分类

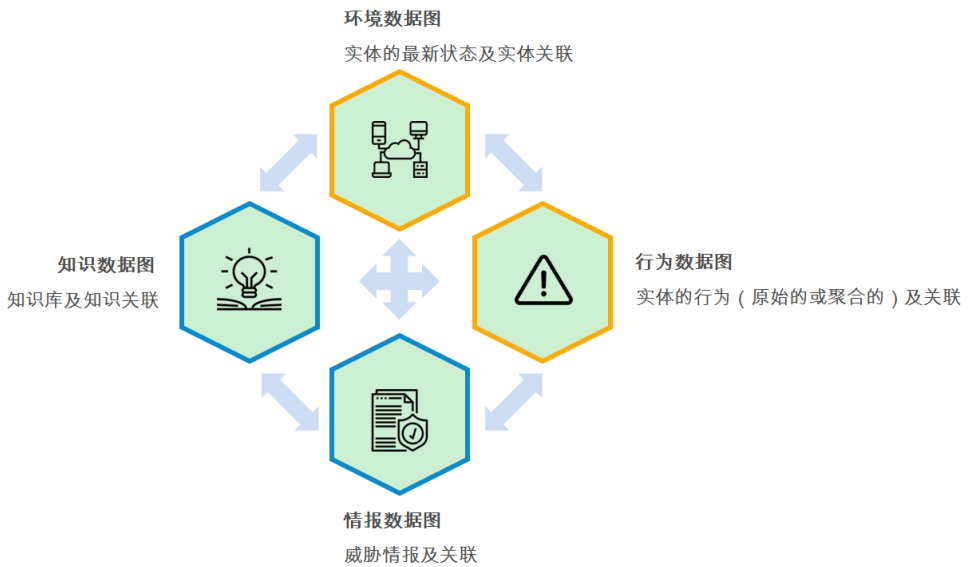


图 10　AISecOps 核心数据图

当前，大规模多维度网络安全大数据的接入，为通过数据分析发现、处置网络威胁带来了全新机会。但考虑到可用的存储、计算资源有限，对安全数据源的甄选和统一处理就显得尤为重要。不同于 DIKW 的数据分层模型 [10] 和 CyGraph 的安全 / 任务知识栈结构 [3]，从网络攻防的对抗本质出发，以给定的网络空间为战场，以保护资产（包括实体资产和虚拟资产）并打击威胁主体为目的，智能化的威胁分析应

该收集并构建以下维度的关键数据图：

- 环境数据图。如资产、资产脆弱性、文件信息、用户信息、IT 系统架构信息等。
- 行为数据图。如网络侧检测告警、终端侧检测告警、文件分析日志、应用日志、蜜罐日志、沙箱日志等。
- 情报数据图。各类外部威胁情报。
- 知识数据图。各类知识库（如 ATT&CK[11]、CAPEC[12]、CWE[13]）等。

各类安全关联数据（包括但不限于以上四个类别）已在很多大数据分析场景中所采用，但仍然没有成熟、统一的体系描述这些数据的分类和使用模式。故应将这里列举的四类数据，从网络威胁事件分析实践出发，通过图结构组织起来，实现每个类别图内关联和不同类别图间关联，以满足网络空间对抗的基本战术需求，包括对环境的掌握、对威胁主体行动的理解、对外部情报的融合以及储备基本知识。四图分立，又通过指定类型的实体进行关联，保证了不同类型图数据表达能力的同时，实现了全局的连接能力。

3.4 AISecOps 技术框架

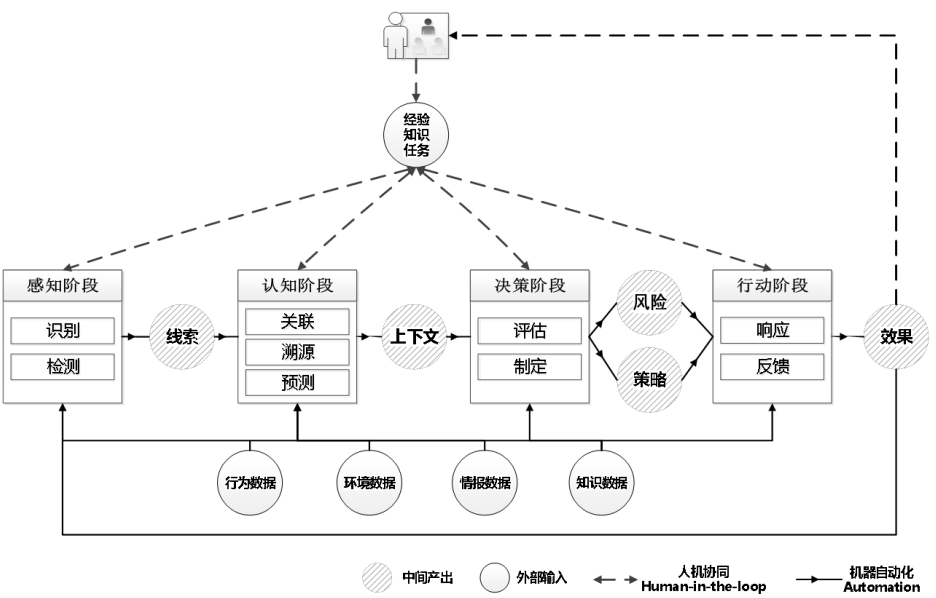


图 11　AISecOps 技术框架

上图阐述了 AISecOps 的技术框架，针对安全运营技术中的关键环节，参考人工智能的经典范式“感知 - 认知 - 决策 - 行动”和经典作战决策 OODA 循环模型的“观测 - 调整 - 决策 - 执行”体系 [14]，进行子任务及其阶段划分，每个阶段包括多个不同子任务。下面详细介绍每个阶段及其完成的子任务。

- **感知阶段** 实现数据级别的融合及基本信息标注，包含识别和检测子任务。识别子任务对大规模数据中的实体（资产、特征、脆弱性等）及其行为的归类、去重、规范化等，以促进多源、异构数据的融合。检测子任务区别于识别子任务，从大规模数据池中捕获、标注异常事件、脆弱性、威胁特征等关键的动静态信息，以标记威胁分析、狩猎、风险分析中的关键线索。
- **认知阶段** 实现线索、事件关联上下文信息的召回与构建，包括关联、溯源、预测子任务。关联子任务通过跨多数据类别、跨长时间周期进行多维信息整合，提供充分的信息连接视图。溯源子任务通过回溯及根因分析，查明、识别事件的起源，明确多事件之间的因果及依赖关系。预测子任务基于当前信息上下文，以路径预测、趋势分析等手段，预判可能的攻击行为、高危的脆弱性等，实现在攻击意图识别和防护方法运用上领先攻击者 。
- **决策阶段** 根据预设目标综合评估风险，实现任务策略的生成，包括评估和制定子任务。评估子任务面向核心运营指标，基于行为、环境、知识等关键信息，持续、综合评估网络安全的整体态势与风险等级，支撑在指定运营成本下的最优研判结果的输出。制定子任务根据动态环境与行为，自适应选择并生成有针对性的、风险驱动的行动计划和策略，明确行动的具体步骤。
- **行动阶段** 根据计划、策略与步骤，协同调动各行动单元完成行动目标，包括响应和反馈子任务。响应子任务完成包括策略下发、设备部署、补丁更新、容错修复等平台级、模块级、设备级、指令集等针对不同层级的风险响应动作。反馈子任务持续收集响应动作执行关联的效果集合，生成面向流程、人、技术多运营要素交互的数据汇总，以支撑自动化任务下一循环的开展。

以上由低到高层次的感知、认知、决策、行动多个阶段及关联子任务是支撑网络安全运营自动化水平提升的关键能力。尽管子任务技术水平可独立演进，但高层次阶段的可用性、鲁棒性仍依赖于低层次阶段技术的成熟度。例如，随着 SOAR（Security Orchestration, Automation and Response）技术的落地，迅速提升了安全运营中各项任务的自动化水平。SOAR 技术提供了以上矩阵中行动阶段响应与反馈的关键能力，提供了数据、流程、技术集成的框架与接口，是 AISecOps 技术实现运营自动化过程中的架构基础实现。现阶段 SOAR 技术已能够协同多模块在多种场景下完成行动阶段响应子任务的自动化，然而在当前威胁溯源、攻击预测、风险评估技术尚未成熟的情况下，自动化响应流程会因误报导致的错误阻

断等动作妨碍系统正常运行，大幅限制了 SOAR 技术的应用场景。

整体上，AISecOps 技术框架包含两个大的循环。一个是图中实线覆盖的机器自循环，这是 AISecOps 追求的运营关键任务自动化的终极目标。另一个是图上虚线覆盖的人 - 机协同循环，这一部分重点描绘人需要参与到运营自动化的每个关键环节中，同时充分获取机器的数据反馈。高水平运营自动化实现的要义仍然是对“数据 - 信息 - 知识”层次化的分析与挖掘，以应对动态不确定性的网络空间环境与高交互的攻防对抗过程。因此，唯有夯实网络空间数据的多层级任务能力基础，才能避免搭建安全任务自动化的“空中楼阁”。实际上，现阶段的威胁识别、溯源、预测等关键技术能力的智能化水平，仍难以有效支持基于 SOAR 的精准响应。事件误判、连接误杀、决策黑箱等多种类型的技术瓶颈，使得更高水平的自动化智能化实现在涉及高风险、关键决策的安全场景下难以有效部署。因此在当前阶段下，人 - 机智能的充分融合，就显得尤为关键了。

3.5 AISecOps 技术成熟度矩阵

安全运营是一项复杂的系统化工程，人工智能技术的应用不是一蹴而就的。如今中心化的安全运营平台收集的流量、日志、标签、指标、事件、规则等多源数据的规模已逾千百倍，智能化、自动化的数据挖掘方法，已得以规模化应用，然而，在数据挖掘与智能分析关键技术的应用实践过程中，研究、开发和运营人员普遍遇到了如数据质量低下、算法拟合粗暴、模型产出易误报、难解释、难维护等问题。

自动化水平	名称	定义	任务阶段									数据交互(DIKW模型)
			感知阶段		认知阶段			决策阶段		行动阶段		
			识别	检测	关联	溯源	预测	评估	制定	响应	反馈	
L0	无自动化	由运营人员全权完成安全运营操作										数据采集
L1	运营辅助	自动化运营系统完成感知、认知、决策中的多个子任务，其他运营操作由人完成										数据集成 信息加工
L2	部分自动化	自动化运营系统针对指定初级任务完成感知、认知、决策、行动全流程子任务，与运营人员进行持续数据交互										信息融合 知识获取
L3	有条件自动化	自动化运营系统完成包含行动层子任务在内的全流程子任务，运营人员须在关键阶段提供适当应答										
L4	高度自动化	在限定场景下，自动化运营系统完成包含行动层子任务在内的全流程子任务，运营人员不一定提供应答										知识理解 知识沉淀
L5	完全自动化	在所有场景下，自动化运营系统完成包含行动层子任务在内的全流程子任务，运营人员不一定提供应答										

完全运营人员完成 → 无运营人员参与

图 12 AISecOps 技术成熟度矩阵

在实践效果难以匹配安全运营迫切需求的背景下，我们提出了分阶段、分层次的 AISecOps 技术成熟度，即自动化能力分级矩阵构建方法，以在统一语义下横纵向定位相关技术的发展层次与现状、应用范畴、应用深度等。

如图 12 所示，按照安全运营关键任务的自动化程度，参考自动驾驶自动化分级 [15]，将 AISecOps 技术的自动化水平划分为 L0~L5 六个层次，对应无自动化到完全自动化。同时，针对安全运营技术中的关键环节，按照人工智能的经典范式“感知 - 认知 - 决策 - 行动”进行概念划分，并基本对应经典作战决策 OODA 循环模型的“观测 - 调整 - 决策 - 执行”体系。其中感知层执行识别（如各类实体识别与分类）和检测（如威胁检测）任务，认知层执行关联（如多源数据集成分析等）、溯源（如回溯攻击路径）和预测（如预判攻击行为）任务，决策层执行评估（如风险综合评估）和制定（如策略、方案生成）任务，行动层执行响应（如部署策略）和反馈（如主动报告）任务。每个任务层次的有效性依赖于上一次层次的成熟度。以下简述 AISecOps 自动化能力的不同级别含义：

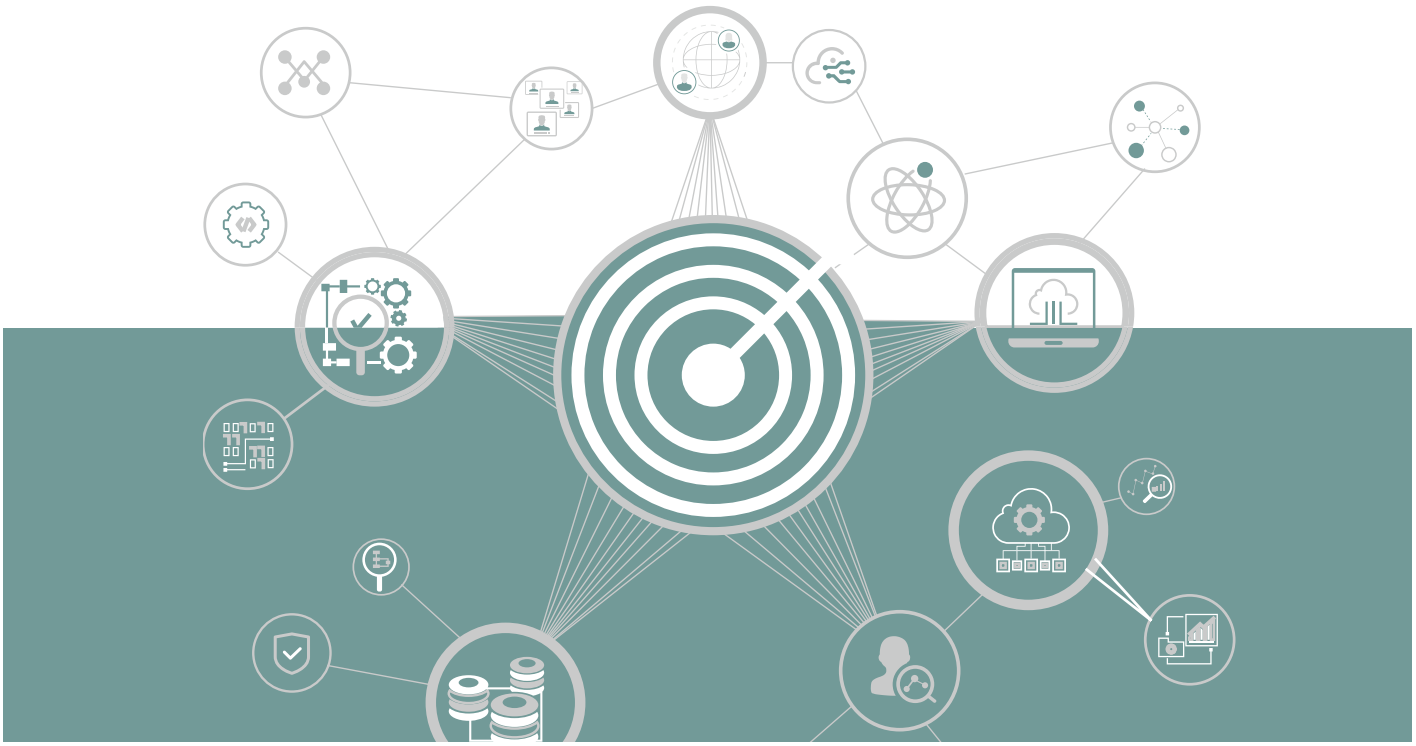
- **L0（无自动化）** 安全运营的所有任务都由安全运营人员完成。尽管 AI 等分析技术能够提供一定层级的识别和检测能力，但该层次下识别与检测不对任何安全运营任务负责，属于较高层级的数据采集能力。
- **L1（运营辅助）** 面向安全运营的运营指标，自动化运营系统开始有针对性的参与环境感知、信息加工认知及风险评估等部分子任务。在该自动化层级下，系统例行数据分析的辅助功能，不参与任何自动化行动子任务。
- **L2（部分自动化）** 针对部分单一环境场景，自动化运营系统参与感知、认知、决策、行动的全流程子任务，与运营人员进行持续数据、知识交互。
- **L3（有条件自动化）** 针对所有任务场景，自动化运营系统完成包括行动子任务在内的所有子任务，但在必要阶段须安全运营人员提供应答与系统接管。
- **L4（高度自动化）** 在限定的复杂场景下，自动化运营系统按照预定的运营指标完全自动化执行，无需安全运营人员介入。
- **L5（完全自动化）** 在所有复杂场景下，自动化运营系统按照预定的运营指标完全自动化执行，无需安全运营人员介入。

通过技术框架的横向技术阶段划分，明确了安全运营技术智能化的关键需求与任务；通过基于技术成熟度的纵向分级，能够有效划定现阶段发展层次与未来的发展方向。以上分类、分级方案，形成了AISecOps 关键能力成熟度矩阵，以期现有的技术方案能够更快速的找到其在 AISecOps 技术领域的定位，并与其他技术能力快速融合互动。

通过 AISecOps 技术成熟度矩阵的构建，能够让技术从业者不囿于技术泡沫造成的困惑。目前来看，在安全运营的智能化技术领域中，我们整体上仍处于 L1~L2 级别的技术发展阶段，多个单点技术水平已经在更高层次有所突破。同时我们所收集的数据、构建的模型、优化的算法及搭建的系统，在特定场景下还未能有效符合安全运营的指标导向性需求，更不用说跨场景、自适应的更高层级运营自动化能力。总之，我们从实践的经验出发，距离高可用的、高自动化水平的智能安全运营技术仍有较远路程。

4

AISecOps前沿技术概述



AISecOps 智能安全运营技术尚处于快速演进的阶段，所采用的技术方案迭代非常快。为了充分探究技术的未来发展方向，定位关键能力瓶颈，本文总结了面向安全运营自动化、智能化的十六种基础前沿技术，并形成技术图谱，以期为网络安全运营场景构建领域技术“内功心法”图谱。

技术图谱在横向上，按照面向攻击对抗的识别粒度进行技术领域划分，粒度自微观到宏观，包括指纹与特征、技术与行为、战术与意图、战役与组织、战役与态势。在纵向上，按照 AISecOps 智能化的经典技术阶段进行划分，包括数据层面的融合建模，以及分析层面的风险感知、因果认知、鲁棒决策、负责行动五大阶段。同时，根据技术的核心数据源不同，通过颜色进行区分，涵盖环境数据、情报数据、知识数据、行为数据以及融合多维的综合数据。通过总结并归类十六种关键技术，试图厘清 AISecOps 的技术分类，以支持技术方案的细粒度抽象与整合，支持安全运营智能技术中台等基础平台能力的构建。

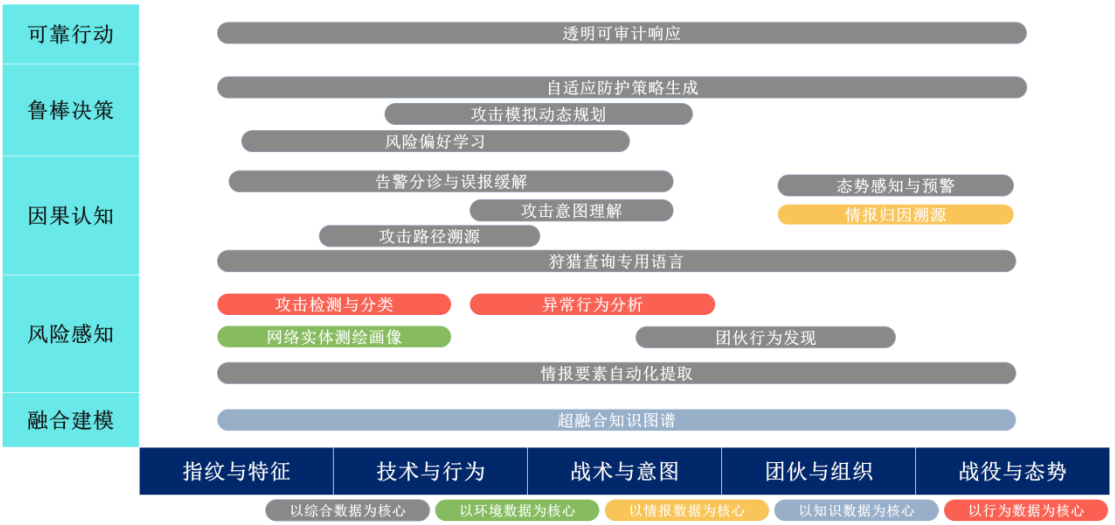


图 13　AISecOps 前沿技术图谱

以上技术图谱中的技术之间有着复杂的依赖关系。整体来看，层次高、位置偏上的技术实现与有效性依赖其下方技术的实现效果。例如，因果认知中的告警分诊与误报缓解技术，依赖于同层次攻击意图理解的建模，以及更低层次的技术，包括风险感知与融合建模的多项子技术。值得注意的是，上图技术的依赖关系与位置关系不是明确对应的。还是以告警分诊与误报缓解技术为例，其与攻击路径溯源技术之间是互相依赖的。溯源技术提供的上下文能支撑更为准确的告警分诊；同时有效的剔除误报、识别高危告警，能够减轻依赖爆炸、降低溯源的难度，提升攻击者、攻击源识别的效率。

以下，首先重点介绍智能化的数据基础——融合建模中的超融合知识图谱技术，进而分阶段介绍其他十五种基础分析技术。

4.1 融合建模

多源、多维、多层次数据的融合建模是后续分析技术的核心基础。本节将重点介绍融合建模中的关键技术趋势——超融合知识图谱。

4.1.1 超融合知识图谱

4.1.1.1 定义内涵

超融合知识图谱的含义是以安全领域知识图谱为核心，面向网络环境数据、威胁行为数据、威胁情报数据、安全知识库等，构建本体化、标准化、全局化的知识结构，支持安全数据的动态查询与聚合分析，提升安全数据运营分析的整体性。超融合知识图谱是后续风险感知、因果认知、鲁邦决策、可靠行动多层次技术能力实现的核心技术基础。没有统一的数据视图支撑，高复杂度算法的构建将是空中楼阁。

4.1.1.2 技术背景

网络环境本身具有典型的图结构，网络安全问题也因此很自然地与图数据结构、图算法结合起来。在 Google 提出知识图谱的概念之后，以知识图谱技术为基础的智能应用方案，已经在推荐系统、问答系统、搜索引擎、社交网络、风控等领域广为使用。在安全领域，最常见的就是各大安全产品中的可视化界面中资产关系图、攻击向量图等。通过图进行数据关联和推理，国内外厂商也在不断地进行深入的尝试。依赖于语义图的内在可解释性，图结构及图算法广泛的应用在诸多场景下，如推荐系统、欺诈检测、网络安全等，为自然存在的大规模数据关系的挖掘提供了系统性的可解释的方法，成为 XAI（eXplainable AI）技术的重要组成部分。此外，针对图算法的研究，如基于深度学习的图嵌入、图遍历、图上异常检测等，增强其可解释性也是重要的研究方向。

4.1.1.3 思路方案

微软的智能安全图（Microsoft Intelligent Security Graph）已几乎全面占领了 Google 引擎“Graph”+“Security”关键词的搜索结果。其通过云生态和平台全面融合，链接多方多维数据，提供全面的威胁关联信息，并以云端的分析能力保证实时的威胁检测，此外还提供了可快速集成的 API。在 2019 年 RSAC 大会上，微软安全团队介绍了数据重力（Data Gravity）的概念，以及云环境下基于

检测和行为图及机器学习的威胁分析算法，该算法能够有效评估事件的风险。Sqrrl（2018 年 1 月被 Amazon 收购）提供网络威胁狩猎平台，结合 UEBA（User and Entity Behavior Analytics）提出过“Behavior Graph”的概念，使用行为评估和关联数据支撑威胁事件的深入调查。发起和构建多个威胁建模知识库（CAPEC、CWE、ATT&CK 等）及相关语言和规范（STIX、TAXII 等[16]）的 MITRE 公司在安全数据的图模型构建方面已有深入的研究。CyGraph 是 MITRE 在图模型研究方面的原型系统。CyGraph 使用了层级的图结构，包括网络基础设施（Network Infrastructure）、安全状态（Security Posture）、网络威胁（Cyber Threats）、任务依赖（Mission Dependencies）四个层次的图数据，用于支持针对关键资产保护的攻击面识别和攻击态势理解等任务。国外使用多源安全数据构建统一分析图结构的项目还有 Cauldron[17]。Cauldron 能够归一化漏洞扫描评估结果，并支持解析多种格式的防火墙规则，通过与网络拓扑的联合分析，能够有效分析网络攻击面的动态变化。

美国的 MITRE 公司研究者提出将任务依赖、网络架构、网络暴露状况以及网络威胁统一组织成多层的知识图谱，通过自定义的图查询语言 CyQL[17]，能够实现诸如威胁狩猎、任务可视化、时序图分析等任务。

ICCS 2018 会议上 IBM 研究员提出威胁情报计算（TIC，Threat Intelligence Computing）的概念，通过构建时序图结构，实现敏捷的网络推理和威胁狩猎。在 TIC 框架下，所有的安全日志、告警日志以及流量日志都存储为统一的时序图，进而通过攻击子图描述威胁或者攻击，威胁发现的问题被转化成子图计算问题。

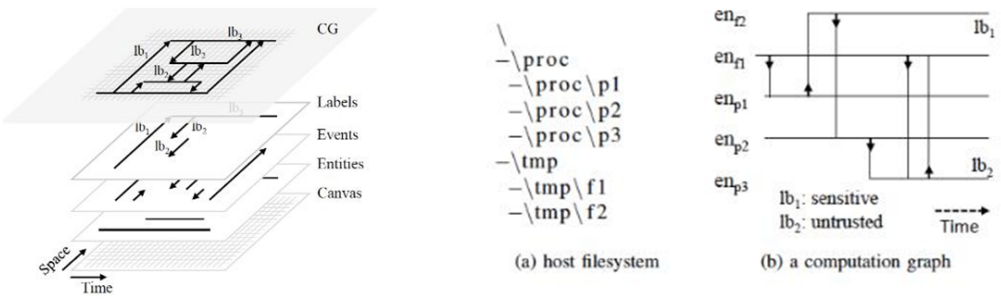


图 14　TIC 模型设计及示例 [18]

国内方面，已有许多产品和研究关注安全数据的图分析方法。例如，绿盟科技结合知识图谱，设计了多个本体对整个网络威胁进行建模分析 [19]，并兼容 MITRE 的 CAPEC、MAEC 和 ATT&CK 等模型的接入与使用，能够从多种威胁情报中提取关键信息并作为知识对知识图谱进行扩展。阿里巴巴利用聚合的

原始告警数据生成有向的攻击图，并通过攻击阶段映射、资产的网络分布及相关边的权重进行告警的优先级评估和攻击场景发现。

4.1.1.4 关键挑战

超融合知识图谱是运营数据关联分析、智能决策、行动响应的重要数据基础设施。尽管近年来有诸多研究工作和厂商产品在持续探索多源数据的融合方案与安全领域知识图谱的构建方法，在超融合知识图谱的设计、技术实现等多个方面，仍存在多方面的挑战。以下介绍相关挑战及技术应对。

本体库设计

图结构设计的一个关键任务，就是设计合理的本体库。本体包括了图中实体（节点）类型、实体的属性类型以及实体间的关系类型（即实体之间边的类型），即表示图结构的抽象概念结构“类”。本体库的构建既要讲科学也要讲艺术。讲科学是指需要遵循一定的规范标准，同时契合适当的威胁模型和描述模型；讲艺术则指的是概念的抽取很多时候是一个仁者见仁，智者见智的过程，并且要符合特定应用场景下的指定需求。例如 ATT&CK 知识库提供了四个核心的实体（战术 , 技术 , 软件 , 组织）及其之间的关系；CAPEC 则主要覆盖 TTP、防护手段、脆弱性等概念；如果直接参照 STIX 2.0，则需要覆盖十余种对象。攻防模拟、威胁狩猎、合规检查、风险评估、检测响应、APT 演练分析等等不同的业务场景，ATT&CK 本身所提供的概念类型是不可能完全覆盖的。因此，ATT&CK 在知识图构建中可作为威胁检测行为模型的知识源和建模方法，而不是一个完备的网络安全知识图。构建可用、可拓展的知识图，在顶层本体结构系统设计的基础上，一方面需要整合吸收所需的公开知识库，另一方面，需要通过知识图谱的手段主动进行知识拓展和延伸。

知识库的关联

以 MITRE 生态下多个知识库为例，包括 CAPEC、CWE、ATT&CK 等，有密切的联系，同时有不同的应用场景。CAPEC 和 ATT&CK 是两种不同的攻击建模方式，CAPEC 针对基于应用脆弱性的攻击，通过攻击模式的抽象和分类，构造了攻击行为的可查询词典；而 ATT&CK 则更贴近威胁检测的实战。

在图 15 中，我们通过 STIX 2.0 架构对比一下两者所处的位置。可以看出两大知识库在概念的表达上有交叉，又各具特点。

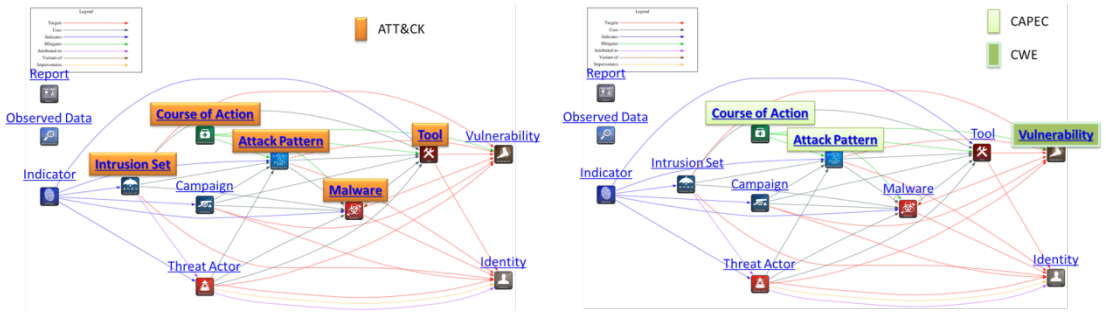


图 15　ATT&CK、CAPEC、CWE 与 STIX 2.0 映射对比

图 16 展示了 ATT&CK 与 CAPEC 攻击模式分类的关联关系。其中 ATT&CK 以战术目标为列组织成矩阵结构，CAPEC 通过攻击模式的抽象组织成树形结构。以 Discovery 战术下的 System Owner/User Discovery 技术为例，与该技术关联的 CAPEC 攻击模式为 Owner Footprinting，同时该攻击模式关联的 CWE 为 Information Exposure。

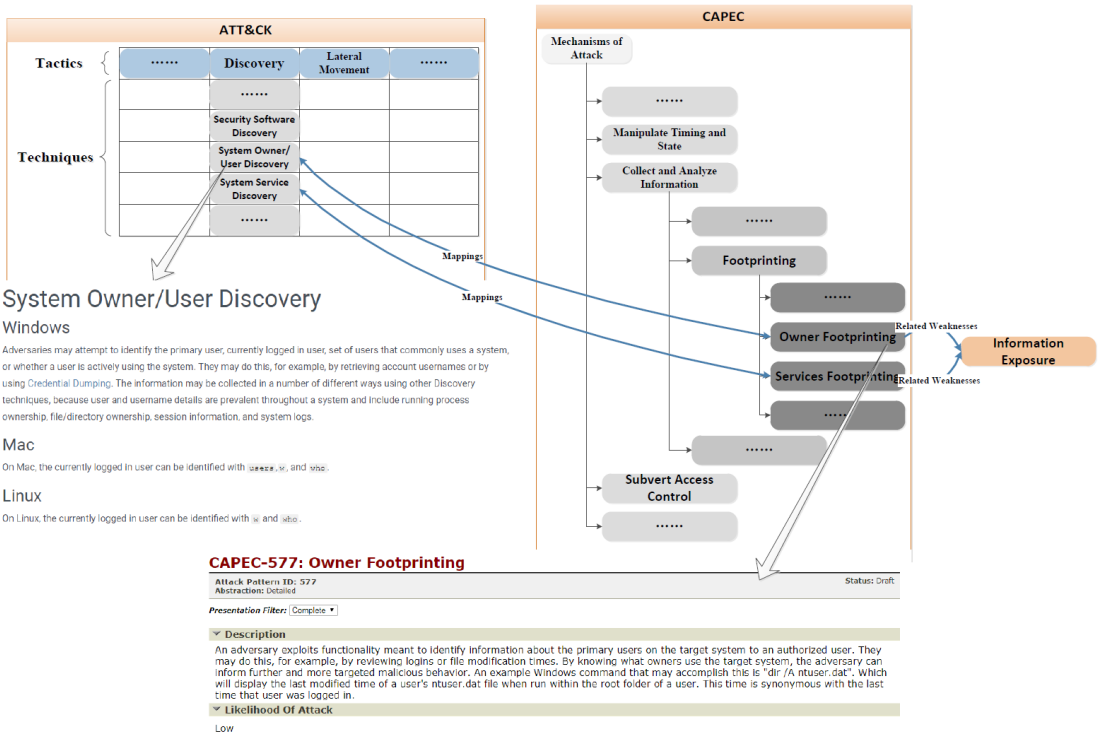


图 16　ATT&CK 与 CAPEC 的映射关系示例

威胁检测的实践不断证明基于行为的检测更能够适应动态环境下的高级威胁分析，不过，特征 + 行为的组合检测能力，是当前威胁检测效率提升的关键。从知识库构建的角度讲，CAPEC+CWE 和 ATT&CK 都是不可或缺的。MITRE 生态的持续完善能够充分降低各个知识库之间建立关联的难度，例如，CAPEC 和 ATT&CK 目前都能够纳入 STIX 2.0 的表达体系；同时，两大知识库之间也已建立了知识的关联引用，当前 ATT&CK Enterprise 对应的 244 个 Attack Pattern 中与 CAPEC 关联的有 44 个。

在威胁建模和知识库积累方面，无论是基于已有的知识库还是通过知识图谱算法抽取知识，构建知识图，一方面需要兼容已有的标准和架构，另一方面，也需要根据实际的应用场景选定合适的知识范围。MITRE 于 2018 年提出过一个针对金融服务机构的增强威胁模型 [20]。该模型虽然采用了较老版本的 ATT&CK 和 CAPEC 知识库，但也为我们展示了两个模型知识库联合使用枚举攻击能力的案例。

威胁模型升级

不同威胁检测方案、设备提供商对威胁事件的理解层次和粒度不一样，输出的事件日志也难以打通。ATT&CK 的出现，为促进统一的知识抽象带来曙光，为提供商自身能力的验证、不同提供商之间检测能力的横向对比、技术能力的共享提供了全新的视角。

无论是基于静态特征特征还是基于机器学习的异常行为检测，各个威胁检测能力提供商往往有自成体系的威胁分析模型和事件命名体系。除非企业方案设计之初即采用了最新的威胁模型，本地化的检测能力要想和 ATT&CK 等知识库进行关联，需要合理的映射机制。很多企业已将 Kill Chain 攻击链模型作为威胁建模的基础，因此转向全新威胁模型体系的过程必然会给整个企业的威胁检测架构带来一定的冲击。威胁模型的升级对相对成熟的安全能力提供商更不友好，因为这些企业往往已具备大规模的 IOC 库、异常行为库，并且对应着各种自定义的命名规范。专家校验和归类自然是必不可少的过程，同时也需要自动化的关联和归类手段。在统一的威胁模型和命名体系下，多源行为图、环境图、情报图才能够有效关联威胁知识图，获取理解行为模式、分析推理的基础知识，打通各类数据间的检索壁垒。

知识抽取与概念对齐

公开的安全领域知识、情报数据一般具有结构化或半结构化的存储、传输形态，以及相对成熟的知识定义标准或规范。然而，在构建场景驱动的领域知识图谱过程中，收集社交网络、文本数据源中的非结构化、未解析的半结构化数据，需要有效的命名实体分析技术、关系推理技术，来辅助提取需要关注的网络空间实体，并对齐到预设的本体库映射关系中。鉴于大规模非结构化文本中包含大量实体和关系噪声，对安全领域知识与情报的抽取，会造成统计层次、语义层次的干扰。一般仍需要特定的语法规则、特征规则，

在知识抽取过程中进行模式和指纹的过滤，以提升抽取知识的质量，提升概念对齐的效率与准确性。

攻击模拟与知识拓展

ATT&CK 矩阵等知识库的构建，不是简单的抽取 APT 情报和相关报告。各种行为的提取依赖的是在特定的场景下复杂、真实网络环境下的攻击模拟与对抗的不断验证、补充、完善。此外，现阶段攻防知识库也远未成熟，针对不同场景、不同领域的威胁行为的知识需要整个社区不断的积累和贡献。因此，将领域共享知识库转化成企业自身的知识图并用于威胁分析，能够提升企业自身的检测能力，但更重要的是需要企业建立自己的攻击模拟环境，验证、精炼、修正知识结构，发现新的知识关联，以适应指定场景下的威胁分析任务。目前，支持 ATT&CK 等知识库模型的攻击模拟或渗透工具已有不少，如 MITRE Caldera[21]，Endgame RTA[22] 等开源项目。搭建攻击模拟环境的要点，基于 ATT&CK 验证流程、设计分析算法以及创建新的 ATT&CK 知识概念，相关经验和手段我们可以通过官方文档深入研究。

4.2 风险感知

感知是智能化技术迈出的第一步。安全运营智能的感知层，主要面向网络空间各类实体与行为的识别与检测，主要包括情报要素提取、网络实体画像、攻击检测与分类、异常行为分析和团伙行为发现等前沿关键技术。

4.2.1 情报要素自动化提取

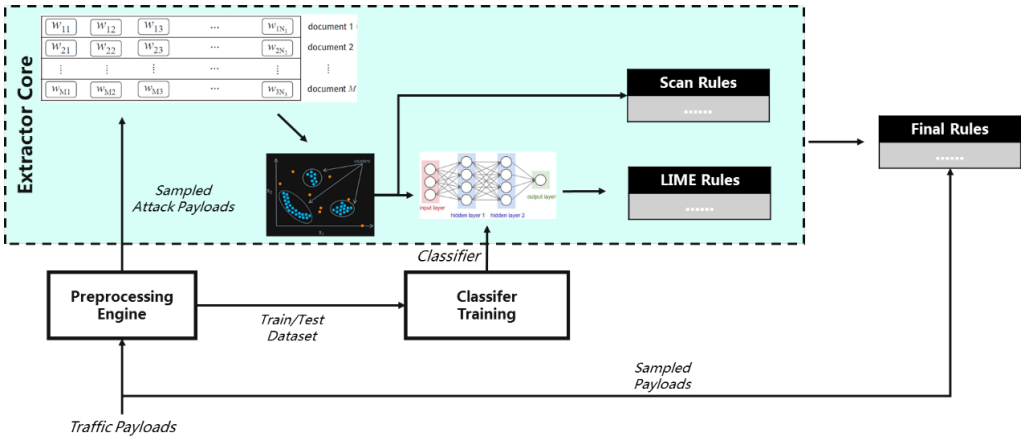


图 17　基于可解释人工智能提取恶意流量特征规则 [23]

4.2.1.1 定义内涵

情报要素自动化提取的含义是通过数据驱动的模式提取方法，从流量、样本、社交网络、情报文本等多源数据中，自动化提取威胁情报要素（攻击者、活动、技战术、特征、防护策略等），支撑网络防御的预防、检测、响应、预测等全周期的信息采集。

4.2.1.2 技术背景

网络威胁情报要素的准确提取关系着网络空间防御的时效性、有效性，决定了网络空间已发生威胁、事件的分析、响应周期，是安全运营流程自动化的重要一环。传统威胁情报的生成依赖事件驱动的专家分析方法，一方面依赖专家经验，耗时耗力；另一方面情报时效性差，情报滞后导致被动挨打。在威胁高度组织化、武器化、规模化的背景下，探索智能化算法与数据驱动的高质量、实时情报要素自动抽取方法，显得尤为关键。

4.2.1.3 思路方案

情报要素自动化提取是一项面向网络安全领域知识构建需求的重要任务，自动化的要素提取，关键技术目标是场景驱动下的模式识别。在攻击特征提取场景下，例如根据模拟的、采集的已知恶意样本、恶意流量，提取恶意特征，经典的处理方法一般可通过传统的序列相似性、文本相似性、结构相似性等手段，快速定位可疑特征信息。此外，基于可解释人工智能方法提取模型的知识，已成为知识获取的重要方法之一，例如通过透明可解释的决策树模型、文本主题模型、图模型、注意力机制等，或黑盒模型叠加后处理（Post-hoc）的解释手段 SHAP、LIME 等等，抽取安全检测分析模型内的攻击模式与特征，如图 17 所示，通过聚类与模型推断算法，能够有效提取恶意文本中的关键词特征形成检测规则 [23]。在攻击组织活动、技战术自动化情报生成的场景下，可通过经典的命名实体识别、关系抽取、知识图谱关系推理等技术手段，提取、对齐、关联情报实体要素，实现情报的标准化与可共享性。自动化的提取方案，能够有效作用在大规模数据空间下，从数据的角度提升威胁特征的区分性、情报实体的全局一致性等。

4.2.1.4 关键挑战

威胁情报的准确性、信息粒度、置信度等指标，是安全运营风险预警、威胁检测、事件处置等各个环节技术实现的关键信息基础。自动化技术的实现需要以情报要素的可用性为核心，提升实时性与处理效率。主要技术挑战包括：

要素特征的语义泛化

传统攻击指纹的提取，在专家经验的优化下，能够在保证安全语义的前提下，提升检测识别的泛化能力。然而，数据驱动的方案普遍缺乏场景语义的限制，统计驱动的、关联驱动的特征规则可能不符合网络或安全的语法语义，导致提取结果的失效。

情报噪声干扰

数据驱动的方法依赖数据的质量，在低信噪比的数据上实现知识提取工程，面临统计失效、语义不明、实体难对齐等多方面的挑战，也对情报数据采集、数据预处理、数据融合等多阶段的情报数据生命周期管理提出跟高的要求。

要素提取的可解释性。

透明度与可解释性是数据驱动方案需要重点关注的技术要素之一。特别是面向安全攸关的威胁情报处理场景，情报的可信度是决定情报源信誉的关键，而情报生成技术的可信任程度，又是情报可信度的重要组成。采用黑盒模型抽取的情报要素，其潜在隐患，包括算法偏见、数据投毒、模型错误等等，可导致情报在实用阶段造成样本误杀、响应失效甚至系统破坏的后果。

4.2.2 网络实体测绘画像

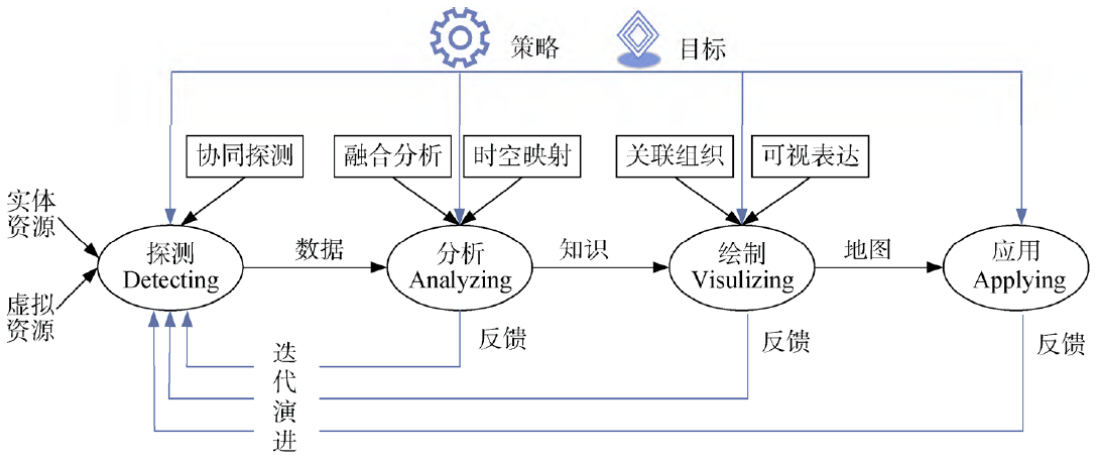


图 18　网络空间资产测绘技术体系 [24]

4.2.2.1 定义内涵

网络实体测绘画像的含义是通过主动指纹探测、被动的信息采集，收集、分析、整合网络空间资产、身份、数据等各类实体及其特征信息，形成网络空间的整体画像和实体局部画像，支持网络风险的全面、深度分析与威胁情报生成。

4.2.2.2 技术背景

随着 IT 系统的多元化、复杂化、高联通，以及网络空间对抗的持续升级，对防守方“知己”的能力提出更高的要求。云计算、5G 移动网络、物联网等跨应用、跨平台、跨基础设施的大规模多样化实体的接入，直接导致传统登记式的、定点监控式的实体的识别、管控方法失效。企业和组织的网络监控盲点，将成为 APT 等高级威胁的突破口，带来巨大的安全隐患。因此，亟需自动化的、智能化的网络实体识别和画像方法，来提升资产等实体管理的自适应性、准确性、实时性，保证安全运营风险管控的覆盖率与精确度。

4.2.2.3 思路方案

网络空间实体测绘的关键是保证实体实例的覆盖率以及准确的动态画像，核心技术主要包含已知类型实体的识别和未知类型实体的分类。已知类型实体的召回，在于通过特征指纹匹配与行为模式匹配，快速召回收录在册的实体类型实例；未知类型实体的分类，需要通过无监督或半监督的特征与行为聚类、信息流或结构性关联分析、统计频繁项挖掘等方法，识别未知实体数据中的模式信息，寻求与已知类型实体的相似性与关联性，并向运营人员提供数据特征支撑人工分类分组标记。值得注意的是，如图 18 所示，网络实体行为及其所处环境的动态性，决定了实体测绘不是一劳永逸的，而是需要持续迭代演进的 [24]。实体探测仅仅是测绘流程的一个步骤，分析、跟踪、可视化已成为实体画像的重要组成。例如，实体画像的准确性决定了基于异常行为分析的 UEBA 等技术方案的成败。

4.2.2.4 关键挑战

针对任何一种网络实体，包括用户身份、各类资产、多形式数据、应用服务、资产脆弱性等等，进行网络空间的测绘都有特定的技术挑战。这其中，测绘画像的共性挑战主要包含：

实体寻址空间爆炸

主动式的实体探测，首先需要按照一定的“寻址”策略来顺序测定实例的网络空间位置。以网络威胁情报的资产、服务识别为例，IPv6 协议带来的地址空间膨胀，直接导致了传统遍历式扫描的失效。

再以面向企业级数据管控的数据资产测绘为例，面对数据类型、数据存储形式、数据访问权限、数据保护措施等因素的多样性，传统规则驱动、关键词驱动的匹配方法不再实用。数据驱动的模式发现与自适应寻址策略是未来网络实体测绘需要克服的难题之一。

实体行为不稳定性、隐匿性

如前所述，网络实体测绘需要持续跟踪与更新，保持实体画像的实时有效性。然而，随着容器化、微服务化等技术的发展，导致资产、服务等实体生命周期剧烈变化；此外，隐私保护等目标驱动下，流量、服务、身份的隐匿性大幅提升。这些都给实体的精确测绘带来前所未有的技术挑战。

4.2.3 攻击检测与分类

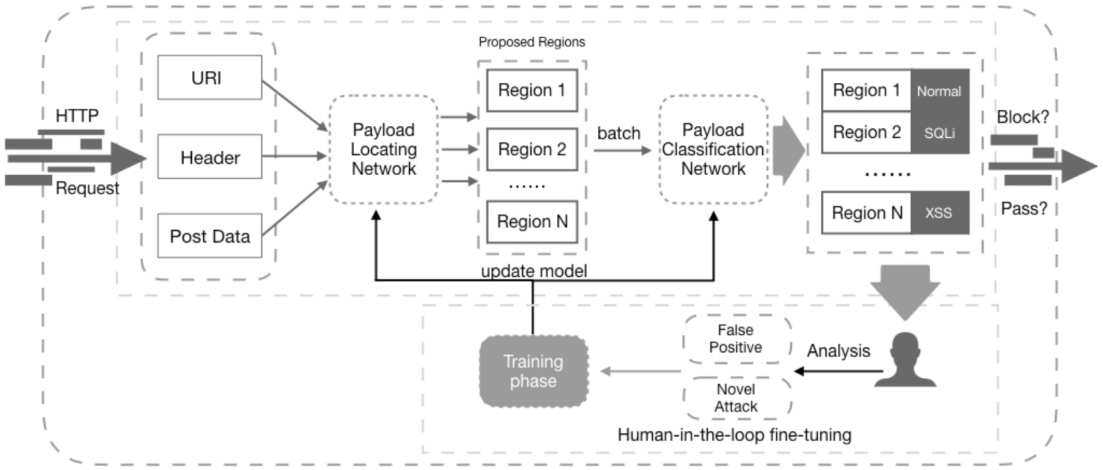


图 19 LTD Web 攻击检测引擎 [25]

4.2.3.1 定义内涵

攻击检测与分类的含义是针对各类网络实体及其行为，通过有监督或半监督学习的方式，实现攻击行为的识别，并区分攻击的技战术类型。

4.2.3.2 技术背景

攻击检测与分类是智能化技术与网络安全数据最早结合的应用场景之一。在入侵检测、Web 攻击检测、恶意样本及其家族分类、恶意流量检测、恶意邮件识别等多种场景中，为了应对爆炸式增长的数据规模及攻击模式，弥补传统专家规则在时效性、准确性、覆盖率上的不足，数据驱动的检测与分类成为

关键补充。

4.2.3.3 思路方案

攻击检测与分类的关键是融合数据特性的算法建模。网络安全领域的算法建模相对于其他产业有一定的后发优势，可根据所处理数据的特性，如事件序列数据、时序数据、文本数据、实体关联图数据等，借鉴相关领域的成熟分析方法与思路。比较经典的方法，有基于集成模型和动静态特征集实现的恶意软件家族分类；基于 CNN+LSTM 和流量数据包、数据流多层次特征的恶意（加密）流量分类；基于图表示学习和进程调用关系的无文件 APT 攻击检测等等，不一而足。如图 19 所示，借鉴目标检测中定位、识别的两阶段识别方法，研究者能够以可解释的方式有效识别不同类型的 Web 攻击 [25]。参考 ATT&CK 模型，现阶段包括终端、网络、文件等多源、多维度的二十余类数据的采集，给威胁分析带来全新的分析机遇。在有效数据标注的基础上，准确的学习攻击样本与正常样本之间的关键模式已不再是难事。

4.2.3.4 关键挑战

检测与分类的关键在于高质量、有标签的数据集。除此之外，建模过程需要应对网络安全数据的分布的不一致性。主要挑战总结如下：

高质量数据标注

这是决定有监督学习领域技术成败的关键因素之一。攻击样本标签化和数据积累一方面依赖研究积累，例如企业在样本研究中的样本分析结果；另一方面，攻击靶场中的自动化攻击模拟能够加速标签数据的收集过程。

训练数据的局限性

攻防博弈持续升级，决定了训练样本空间只能覆盖有限已知攻击类型的已知实现手段。这种空间分布不一致性导致训练模型上线后分类和检测性能迅速衰减的困境。

复杂的数据编码、混淆、加密

攻击者的高对抗性，例如自定义数据编码、对抗性混淆、隐匿通信等等，体现在数据上是难解析、难识别、难定位。许多经典智能算法无法直接应用于安全数据挖掘。无视安全语义的、端到端的统计学习模型已被证明无法有效应用在安全场景下。

4.2.4 异常行为分析

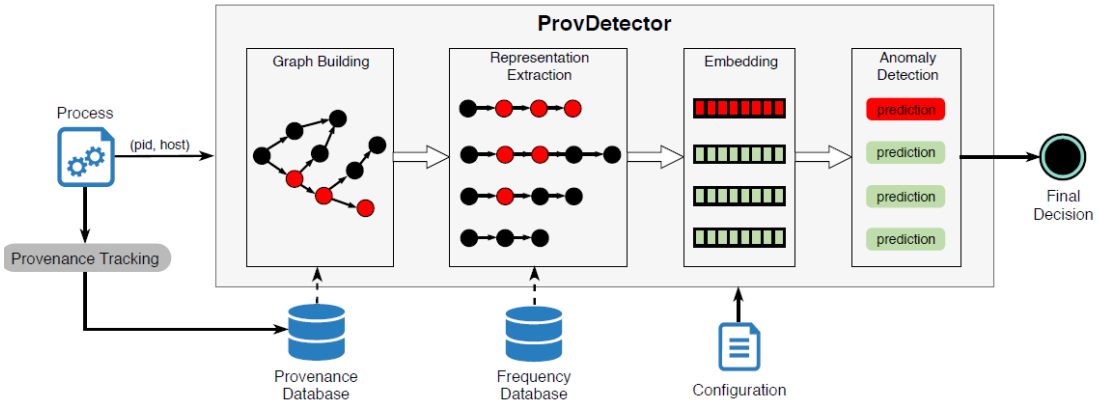


图 20 终端进程行为异常行为检测 [26]

4.2.4.1 定义内涵

异常行为分析的含义是构建多层次网络实体的行为画像，识别偏离正常行为基线的行为模式，捕获、召回潜在威胁线索与攻击行为。

4.2.4.2 技术背景

威胁狩猎的关键任务之一是发现威胁线索，从线索入手顺藤摸瓜，以识别攻击者攻击意图，还原、预测完整的攻击者行为序列。然而，高级威胁具有低频性、隐匿性、对抗性，无文件攻击、隐蔽信道等攻击技术已成为攻击者突破防御措施的重要手段。然而传统静态指纹规则驱动的检测方法依赖专家经验，有监督的威胁检测与分类方法对训练数据的采集、标注要求较为苛刻，以上两种经典方案主要针对已知威胁检测，同时在动态对抗环境下检测效果不理想。

4.2.4.3 思路方案

异常行为检测的关键是正常行为模式建模与离群（异常）点检测算法设计。行为分析的主体是网络环境下的各类实体，包括系统相关的（进程、网络、文件等）、应用相关的（API 调用、业务数据流等）、用户相关的（登录、访问等）等多维度、多层次可观测数据源。针对任何一类实体行为数据的建模，可对应一种具体的威胁分析场景。行为画像建模的关键在于通过统计建模、机器学习、策略抽象的方式，识别实体正常行为的关键参数与结构。常用的技术包括频率统计、聚类、编解码器、时序模型、隐马尔

科夫建模等。在行为画像模型的基础上，对动态输入的未知行为执行离群点检测。离群点或异常点，指在数据模式中与大多数据点特征偏离较远的点。离群点的检测技术实现基于行为画像模型的构建方式。从数据特征建模的角度来看，主要包括基于距离的方法、基于密度的方法、基于统计阈值的方法、基于信息熵的方法、基于图的方法等等。不同的场景下，异常行为分析的数据粒度可能不同，整体来看，行为分析具有较强的环境自适应性，并且不依赖特征指纹与恶意样本，能够有效召回不同网络环境、不同攻防周期内的异常行为，是对传统静态的、针对已知威胁检测的有效补充。

用户及实体行为分析 UEBA 技术的核心实现就是异常行为分析。行为分析是分析攻击者战术、技术、过程（Tactics, Techniques, and Procedures, TTPs）的基础，是一个动态分析过程。分析中实体（包括 UEBA 中的用户与实体）的选择，决定了分析场景与数据采集的粒度。以图 20 为例，通过异常进程行为分析技术来实现对 ATT&CK 矩阵中的 T1055 进程注入（Process Injection）这一攻击技术的检测 [26]。该方案通过采集终端侧的进程行为数据，构建进程调用依赖图（Process Provenance Graph）。基于异常共现频率分数实现图上异常路径特征提取，进而通过语言模型完成进程实体序列的向量化表示，实现进程异常派生关系的识别与检测。该方案不依赖任何进程黑白名单、进程调用特征，仅通过进程的历史调用模式完成隐匿恶意文件的动态行为分析。

4.2.4.4 关键挑战

异常行为分析是典型的数据驱动威胁狩猎方法。不同于规则驱动和样本驱动的检测，异常检测方法对数据有很强的敏感性，并且整体缺乏安全语义支撑，异常检测结果误报率较高。以下总结技术实现的关键挑战。

行为画像的鲁棒性

网络空间数据的画像面临数据噪声难识别、行为模式混杂、缺乏稳定性等多因素的技术难题。然而，异常检测的关键在于对正常模式的精确建模，所采用的模型、算法缺乏鲁棒性将导致后续的异常检测性能大幅衰减，误报率大幅提升。

异常检测的可解释性

离群点、异常点检测方法普遍基于高维度数据特征。抽象特征、黑盒模型所产生的异常判断难以被运营人员所理解，将限制异常行为分析结果的可信度、可用性。

阈值判断的自适应性

影响误报率的另一关键因素，是异常检测模型的阈值设置。当前，离群点的检测仍然依赖关于“偏离阈值”的先验知识。偏离阈值可以是距离、密度、关联度等与数据结构、模型算法相关的经验参数。宽松的阈值设置会导致系统触发大量非威胁相关的误报，严苛的阈值可能导致攻击事件漏报。根据环境动态变化自适应、自动化调整参数阈值，才能合理平衡指定场景下的误报代价与漏报风险。

4.2.5 团伙行为发现

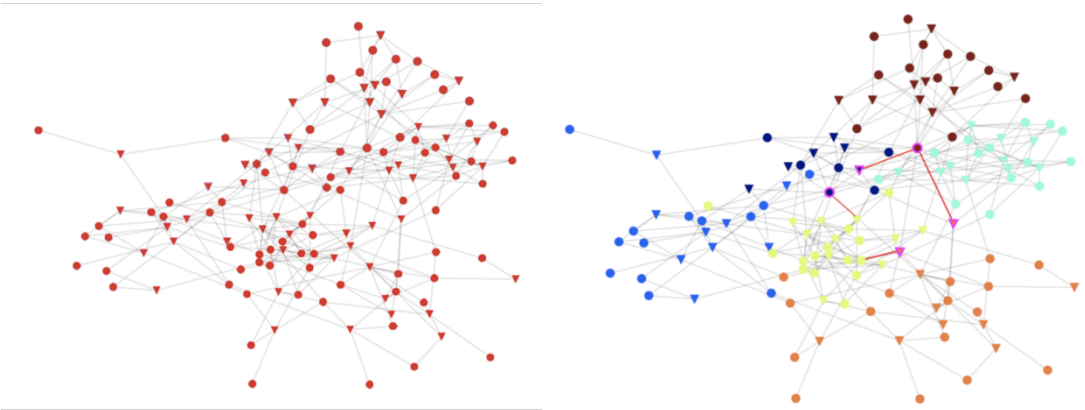


图 21 认证和访问行为社区发现 [27]

4.2.5.1 定义内涵

团伙行为发现的含义是跨时间周期、跨阶段提取攻击、行为事件的行为模式，通过社区挖掘等方法实现攻击者组织、团伙的定位和划定，进而实现对相关事件的归因和追踪。

4.2.5.2 技术背景

网络空间攻击向着组织化、规模化、武器化、服务化的方向持续演化。利益驱动下的攻击组织、团伙控制相对稳定的攻击资源，形成对指定目标的精细化打击。特别是在威胁情报领域，唯有有效定位、识别攻击组织团伙，才能持续感知攻击方的资源调度与武器演化，以实现对攻击源的定位，以及潜在威胁的预防。

4.2.5.3 思路方案

攻击团伙发现的关键是基于威胁数据生成关联图与图上社区发现。STIX 是 MITRE 发起的威胁情报

交换语言和标准，在 STIX 2.0 体系的促进下，全球威胁情报的共享、关联开销大幅降低。通过情报数据图的实例化网络图构建，攻击者、IOCs、技战术、恶意软件、攻击战役及攻击组织等实体及其行为关联能够统一在一张数据图之中。同时，通过语义规则、统计规则、特征命中等方法，对图上的实体点和关系边进行特征抽取，以支撑图结构关联之上的细粒度分析。进而，针对情报的数据规模大、点边特征维度多、置信度差异大等特性，一般采用图社区发现算法实现自动化的团伙标定。社区发现的常用技术包括基于模块度优化的方法、基于谱分析的方法、基于信息论的方法、基于标签传播的方法及基于深度学习的方法等等。数据驱动的攻击团伙发现是一种情报或行为数据增强技术，基于动态情报数据的结构关联性、特征关联，召回疑似团伙、组织，并刻画其行为模式，有助于完善攻击事件的证据链，提升情报置信度。

如图 21 所示，通过构建认证和访问事件行为数据图，在关联图上使用 Louvain 社区发现算法识别图结构层次的用户、服务设备等实体的网络社区 [27]。进而，识别定位关键的跨团伙社区的访问路径，来辅助动态策略部署。

4.2.5.4 关键挑战

攻击团伙发现基于经典的图模式挖掘算法，是情报数据精炼提取的关键环节之一。虽然基于图的关联分析有较强的可解释性，易于运营专家理解，但在情报质量、算法实现、效果验证等多个方面存在技术瓶颈。

情报质量评估

网络空间威胁情报数据呈现爆炸式增长，在迎来威胁信息共享时代来临的同时，情报质量的管控缺稍显滞后。在网络对抗的大背景下，情报信息量、情报时效性、情报真实性等因素可导致情报失效甚至情报污染。在低质情报上构建的分析资源将导致错误的情报推断结果。

统计关联弱相关

攻击团伙的分析不限于图结构的强关联关系，还包括通过特征相似性构建的统计弱相关性。弱相关性的引入一方面是对情报信息的富化，扩展了情报应用的视角。另一方面，数据本身所含的噪声以及模型拟合过程的误差，将引入假相关性，这对依赖高置信度情报的应用场景是不可接受的。

团伙证据验证

攻击团伙发现大部分情况下是无标签样本的分析任务，所得团伙标签和行为模式存在难以验证的挑

战。因此，需要多维、多源情报，以及企业或组织内部其他数据源的旁证策略，如事件溯源机制、主动诱捕系统，来支持算法层级的优化和迭代。

4.3 因果认知

认知是智能计算的核心环节。从网络安全的场景出发，认知层需要解决事件的因果关系的建模问题，以提供支撑后续决策的上下文信息。主要包括狩猎查询专用语言、攻击意图理解、攻击重构溯源、威胁情报归因、告警分诊与误报缓解和态势感知与预警等前沿关键技术。

4.3.1 狩猎查询专用语言

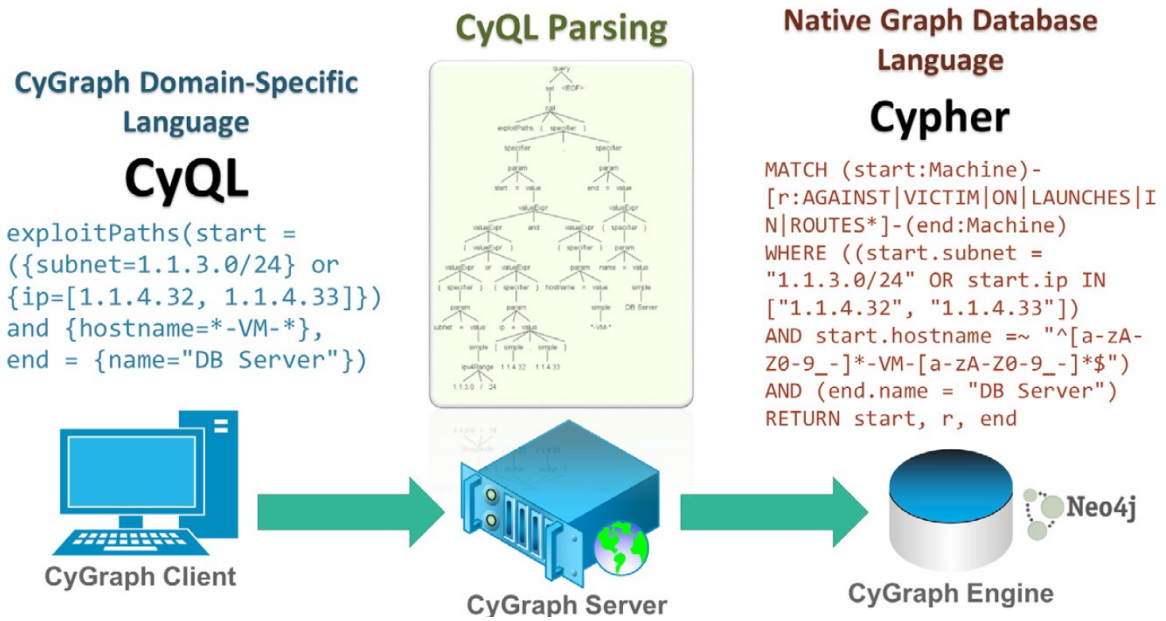


图 22 CyGraph DSL（领域专用语言）[3]

4.3.1.1 定义内涵

狩猎查询专用语言的含义是面向安全运营威胁狩猎已知信息的高效检索需求，基于融合的情报、行为、环境、知识数据基础，设计满足实时性、完整性、准确性的数据检索语言及处理引擎，支撑线索的定位、事件关联信息的召回、情报与知识的准确定位等任务。

4.3.1.2 技术背景

安全运营大数据的存储与检索平台，是支撑运营事件快速推理、响应、报告的重要技术设施。数据的检索性能，决定了情报关联、事件溯源、脆弱性定位、策略选择等环节时效性的上限。现阶段，包括结构化数据库、非结构化数据库等类型存储基础设施已成为智能安全分析平台的主要组成部分。虽然基于多种类型数据库的数据检索，例如针对图数据库、全文检索数据库等，都有特定的、成熟的检索方案与语言，但目前仍然缺乏面向运营检索效率提升的、支持威胁语义的、异构数据库关联的一体化检索方案。这其中最关键的技术环节，就是针对威胁狩猎场景的领域专用语言（Domain Specified Language，DSL）设计。

4.3.1.3 思路方案

威胁狩猎专用查询语言的设计的关键在于业务驱动的定制语义、语法以及支撑结果查询的匹配算法。语义、语法的设计的驱动力是威胁狩猎的关键场景，需要支撑包括不同数据源（如外部威胁情报、内部关键线索等）以及不同模式（精确匹配与模式匹配）的组合查询问题。DSL 一般是声明式的独立抽象层，安全运营场景下最直接的构建基础是融合的知识图框架。基于安全领域知识图图谱，结合其本体化设计与层次化实体交互行为，设计针对指定任务的抽象查询语法。经典的语言设计方案包括如图 22 所示的基于 Cygraph 的 CyQL （CyGraph Query Language）[3]、IBM 的 τ-calculus 等。在匹配算法方面，一方面可直接将 DSL 直接编译为底层数据库查询语言，直接调用数据库内置匹配算法进行数据查询；另一方面，可通过子图对齐与相似性匹配、图神经网络、表示学习等方法，基于分析算法，从大规模数据中查询攻击模式、关联线索。

4.3.1.4 关键挑战

威胁狩猎是主动防御的重要环节，根据线索，快速、准确的关联信息查询，是提升狩猎效率的关键。DSL 的设计既有科学又有艺术性，针对网络安全场景，主要挑战包括：

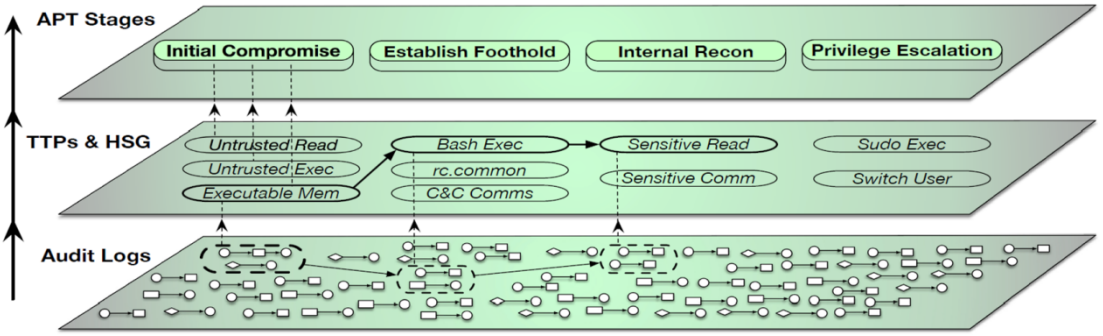
查询语言的灵活性

图数据库已成为威胁狩猎领域的数据库新宠。目前，简单的利用图数据及其内置查询语言进行威胁信息的定位，一方面数据模式过于扁平，难以满足威胁检测、模式识别、意图抽象等多层次的不同查询目的；另一方面，强于结构关系查询，弱于时序依赖查询。这两方面是构建更灵活的威胁狩猎查询专用语言有待解决的关键问题。

分析的效率与准确性

数据查询的底层模式匹配算法，是查询结果有效性的关键基础。底层数据库成熟的匹配算法之外，针对高层次威胁狩猎任务的实际需求，越来越多的分析模块集成到数据库之中，包括在线表示学习、定制化相似度计算与路径搜索、可解释图神经网络等等，这些模块在提升分析准确性、多样性的同时，给传统的数据查询任务带来更大的计算开销。分析能力的实现需要充分根据业务场景优化分析算法与分析架构，以满足最基本的查询实时性要求。

4.3.2 攻击意图理解



APT Stage	TTP	Event Family	Events	Severity	Prerequisites
$Initial\_Compromise(P)$	$Untrusted\_Read(S,P)$	READ	FileIoRead (Windows), read/pread/readv/preadv (Linux,BSD)	L	$S.ip \notin$ {Trusted_IP_Addresses}
	$Make\_Mem\_Exec(P,M)$	MPROTECT	VirtualAlloc (Windows), mprotect (Linux,BSD)	M	$PROT\_EXEC\$ \in M.flags \wedge \exists\, Untrusted\_Read(?,P') : path\_factor(P',P) <= path\_thres$
$Establish\_Foothold(P)$	$Shell\_Exec(F,P)$	EXEC	ProcessStart (Windows), execve/fexecve (Linux,BSD)	M	$F.path \in$ {Command_Line_Utilities} $\wedge \exists\, Initial\_Compromise(P') : path\_factor(P',P) <= path\_thres$

图 23　APT 行为的技战术理解 [28]

4.3.2.1 定义内涵

攻击意图理解的含义是基于大规模、依赖复杂、跨长时间周期的原始日志、检测日志等基本数据线索，从能力水平、攻击阶段、攻击目标等角度，提取、标注、归纳攻击者的战术意图，以明确线索之间的高层次逻辑关联，跟踪、预测攻击者的行为。

4.3.2.2 技术背景

各类检测设备以不同的检测视角与方法，所生成的线索数据，能够捕获真实攻击者的行为踪迹，同

时也无法回避的引入误报与低质量信息。整体来看，攻击者的行为步骤是蕴含潜在方法论支撑与目标导向的。因此，理解和整合多源的、多层次的威胁线索，推测攻击者的攻击意图，减少无关线索对事件分析的干扰，已成为情报关联、行为关联驱动下运营平台智能化的关键能力。

4.3.2.3 思路方案

攻击意图理解的关键在于数据的安全语义化。即通过对数据及其特征的模板化、标签化、体系化归并，形成预设威胁模型框架下的实例化表达。核心技术实现一方面是数据的归一化化与规范化清洗；另一方面，是语义抽象算法，主要可分为两类：基于行为模板的和基于统计切分的。基于行为模板的方法示例如图 23 所示，HOLMES 系统通过预设的数据模式提取策略，将终端侧溯源数据图中的关联日志实体和关系进行抽取，形成符合 ATT&CK 矩阵模型的技战术高层关联图谱 [28]。基于统计切分的方法，通过日志实体的逻辑关联或时序关联，在图数据或序列数据上应用社团发现、标签传播、主题模型、情感分析等经典技术手段，对图上或序列数据进行统计切分和聚类，再结合专家经验的标签化过程，形成符合威胁语义模型的数据基础。

4.3.2.4 关键挑战

攻击意图理解的难点在于如何对齐数据特征与安全语义。尽管越来越多的方案开始注意到意图归纳的重要性，但有效的意图提取技术需要克服以下挑战：

对精准数据标注的依赖

无论是基于统计切分的还是行为模板的语义抽象技术，都离不开专家先验知识的标定。特别是行为模板方案，为了有效限制了数据归纳过程的发散性，依赖细粒度的文件敏感性、行为可信度等标签。这些数据标记过程，一方面需要自动化手段的支持，例如敏感文件自动化识别等，另一方面需要专家的参与，这些都增加了技术实现的难度。

技战术的一词多义

以 ATT&CK 为例，一个技术可能横跨多个战术实现，并以不同的粒度出现在一定的威胁上下文中。因此，需要通过合理的建模方法，识别在不同上下文环境下的不同事件意图，以合理归并、梳理事件的关联关系，厘清事件依赖，发现攻击者的技战术思路与攻击目标。

4.3.3 攻击路径溯源

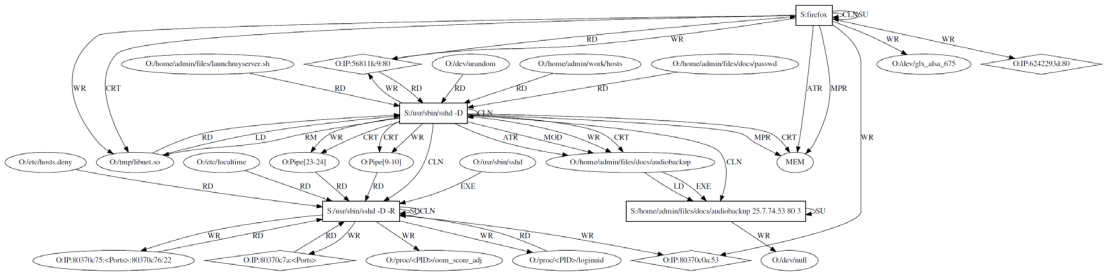


图 24 APT 攻击路径溯源 [29]

4.3.3.1 定义内涵

攻击路径溯源的含义是基于关键威胁线索，结合动态行为与资产环境，融合终端、网络、脆弱性、威胁情报等多源历史日志，回溯、精炼、重构攻击者的行为数据流，完整呈现、还原日志级别细粒度的攻击过程及攻击结果，支持事件调查与取证。

4.3.3.2 技术背景

XDR、SOAR 等解决方案的技术关键，在于融合所有可能的关联数据，形成浓缩的、可运营的事件及上下文，准确定位攻击行为在采集数据上的映射。以终端侧数据为例，溯源数据（Provenance）能够忠实记录终端上实体的行为逻辑依赖关系，自然形成溯源数据图（Provenance Graph，简称溯源图）。所记录的实体，包括文件、网络、进程等维度；根据实体对的类型，实体间关系又包括文件读写、进程创建、网络连接等等。在溯源数据完整有效采集的情况下，通过溯源图的后向追溯（Backward-trace）和前向追溯（Forward-trace），能够有效弥补网络侧的数据盲点，实现攻击事件的溯源与取证。然而，由于缺乏高效、内生的信息流跟踪机制，准确的提取、构建完整攻击路径，仍需要数据的动态分析机制辅助。

4.3.3.3 思路方案

溯源重构的技术基础，是刻画、跟踪行为信息流，以指定的攻击树、攻击图等形式组织相关日志，形成事件前因后果。如图 24 所示，是基于终端日志数据进行攻击溯源与 APT 事件重构的示例 [29]。从数据的角度来看，可将溯源过程建模为统计相关模型、信息传播模型、图关联模型、因果模型等。统计相关性建模主要通过频繁项 / 模式挖掘、注意力机制驱动的序列模型等方式，识别统计层面的实体与行为

关联性，以定位与关键线索相关的最可疑证据链。信息传播模型，基于图数据和标签传播，或先验传播策略，主动跟踪关键操作、敏感数据的传播路径。图关联模型，同样基于图数据，通过图神经网络、可解释图模型等模型算法，识别、抽象可疑的实体与子图结构，以及实体、子图之间的关键行为边，从而实现全局的攻击事件高效抽取。因果模型，相对经典统计模型主要考虑数据的相关性，因果建模通过因果推断框架，如基于约束的贝叶斯网络、反事实推理等，构建具有相对稳定性结构的数据因果依赖链路与图，以探索所采集各类传感器数据间的派生模式。整体来看，溯源与重建的关键在于数据的确定性关系推理。

4.3.3.4 关键挑战

溯源结果能够作为威胁狩猎的关键资源，为威胁的分诊、评估、取证提供丰富的上下文。不过，仍然没有免费的午餐。攻击路径溯源技术有着多方面的挑战，以下简要分析。

溯源图依赖爆炸

这是溯源数据分析中的个性化问题。还是以细粒度的终端数据为例，根据采集方式，溯源数据可分为两类：细粒度的（Fine-Grained）和粗粒度的（Coarse-Grained）。因现阶段性能和系统架构易用性限制，粗粒度的溯源数据广泛应用和部署。粗粒度的采集采取“贪心”的方式，记录实体间所有可能的依赖关系，难以准确跟踪实体间的信息流向。即，某个实体的下游实体的信息流，可能由时间较早任意一个实体信息流产生。特别是长期存活实体的存在，这种不确定性会造成上下游实体的信息依赖的爆炸式增长。

关键线索的缺失

攻击者的高对抗性、采集系统的欠稳定性，都可能导致数据层次关键日志线索的丢失。在证据链、行为序列断裂的情况下，需要鲁棒的分析算法支撑事件重建，包括知识推理算法等关系推测及补全技术需要针对网络空间数据进行优化和适配。

性能拓展性瓶颈

威胁狩猎可包含如已知威胁实时匹配的 OLTP（On-Line Transaction Processing）任务及长周期、大规模关联分析及溯源的 OLAP (On-line Analytical Processing) 任务。为保证不同任务特别是 OLAP 任务下数据的可用性，溯源数据规模将迅速膨胀。此外，终端的多样性，将在数据生命周期、数据对齐、数据关联等多方面带来存储、分析架构的拓展性冲击，实现完整的攻击事件取证还原，同时保持系统的高性能将充满挑战。

4.3.4 威胁情报归因

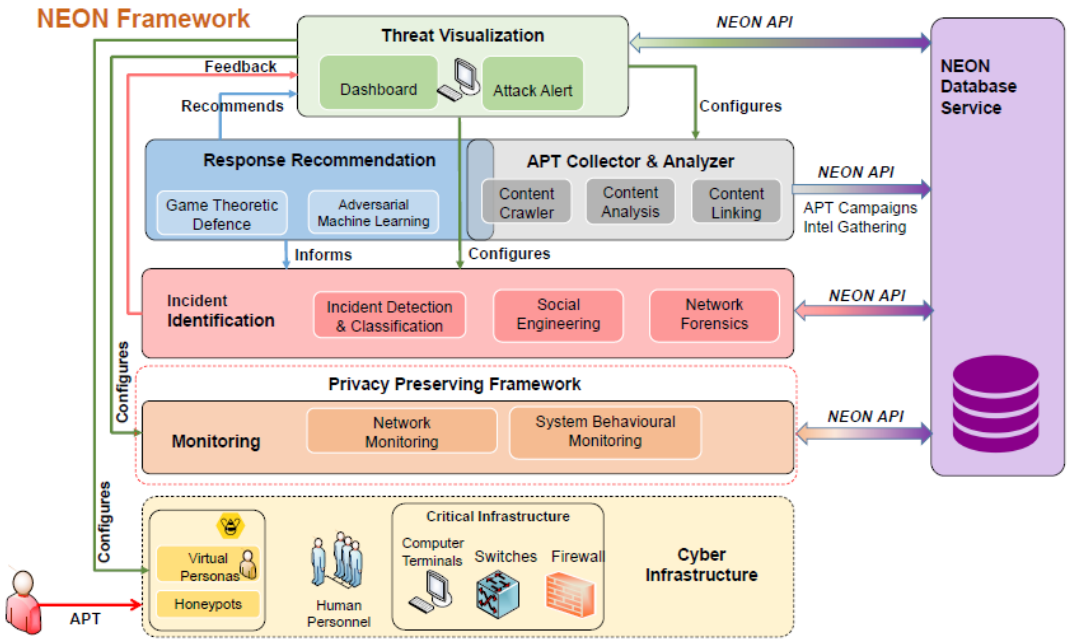


图 25 增强网络攻击归因框架 [30]

4.3.4.1 定义内涵

威胁情报归因（Threat Intelligence Attribution）的含义是基于威胁情报中的关键要素，例如技战术模式、攻击基础设施、恶意软件基因、攻击意图与目标等，突破攻击行为伪装，识别、定位特定的攻击者、攻击组织等威胁主体，为事件的取证、溯源、归因提供基础，为防御反制措施的实施提供高置信度证据支持。

4.3.4.2 技术背景

APT 等外部高级威胁行为，具有高破坏性、对抗性和隐匿性。对攻击个人与组织的追踪与定位，是攻击行为、攻击活动分析的关键环节。在网络安全运营工作流中，外部威胁源的定位与跟踪，将直接关系到企业和组织整体风险策略的制定，例如威胁情报订阅、防御策略生成机制、策略执行置信度、事件报告结构完整性等多方面，是运营能力的重要组成。因此，基于威胁情报的归因溯源技术已成为网络安全运营技术的必行方向之一。

4.3.4.3 思路方案

基于威胁情报实现攻击行为、事件归因的关键，在于情报的深度关联与置信度评估。在情报深度关联方面，最重要的驱动力还是情报的标准化与规范化。这一点上 STIX 2.0 情报标准、ATT&CK 技战术矩阵、CAPEC 攻击和脆弱性枚举库等开源数据库、标准的完善，推进了整个网络空间威胁情报体系水平交互的完备化。此外，情报与本地化分析检测数据的联动，是情报细粒度语义富化等垂直交互的重要组成。经典的数据驱动情报关联方法包括基于草图提取（Graph Sketches）的情报聚类方法、基于子图模式搜索的情报行为匹配、基于基因 / 血缘分析的恶意样本关联、基于知识图谱的语义推理关联等。情报关联之外，威胁归因的关键在于提升情报数据的置信度。置信度的评价一般通过基于区块链的情报信誉机制、基于证据关联命中评级方法、基于情报数据共享多方计算融合等方式实现。整体来看，威胁情报归因的可用性首先是机制保障驱动的，并通过数据智能支持证据强化。如图 25 所示，研究者提出了增强网络攻击归因框架 [30]，该框架融合了多层次主动、被动的攻击检测与对抗技术，并构建人机协同的分析闭环，以有效定位攻击事件所关联的攻击组织。

4.3.4.4 关键挑战

基于威胁情报的攻击组织归因，最本质的挑战包含两个方面，一个是数据层次的融合问题；另一个是由攻防的高度对抗性导致的可信度问题。具体表现在：

攻击者的高度对抗性

攻击的组织化、产业化，在恶意软件定制化、攻击代理池化、网络连接黑箱化、身份窃取混淆等多个维度，为攻击者身份的识别和定位带来巨大挑战，目前尚未有单一自动化技术方案，能够产生证据链丰富完整的归因溯源成果。因此融合蜜罐、黑产情报、企业内部线索等多方面数据并在关键实体与关系上进行对齐融合是归因分析的基础。

负责的归因结论

归因结论的置信度直接关系到策略执行、防御反制、损失定责等多方面的技术联动。尤其是针对组织级、国家级关键信息基础设施的攻击归因，需要精确定位、证据确凿，以有效支撑高层次决策。归因结论的鲁棒性依赖负责任的数据智能，目前仍然存在策略偏见、透明度低、因果性差等多方面的技术挑战。

4.3.5 告警分诊与误报缓解

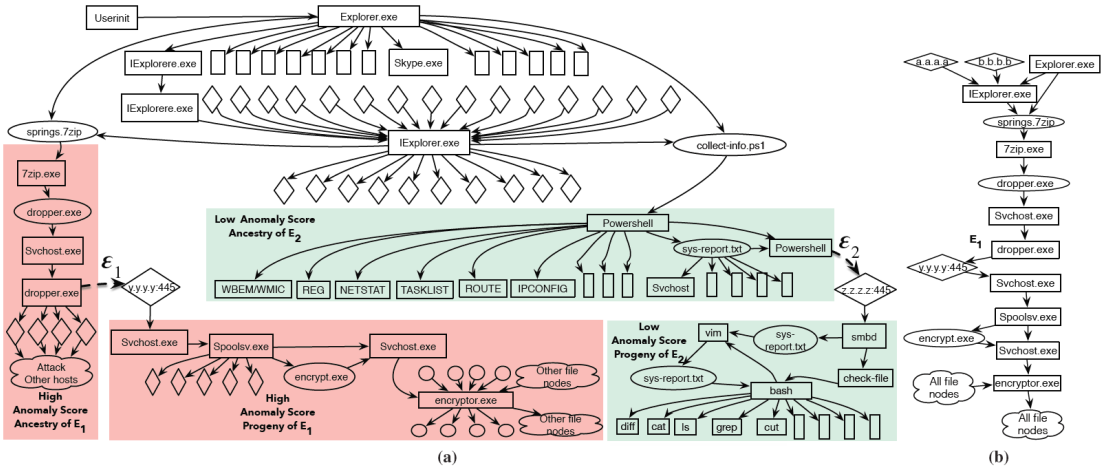


图 26 抗误报疲劳的告警分诊 [31]

4.3.5.1 定义内涵

告警分诊（Alert Triage）与误报缓解的含义是基于告警统计、时序、语义、关联等维度上下文，对告警进行自动化分类，并评估其威胁等级，向运营者提供基于风险的告警排序列表，降低误报对事件调查的干扰。

4.3.5.2 技术背景

随着防御方检测能力深度和广度的增加，安全运营中心被动接入的告警规模如今已膨胀百倍千倍不止。安全运营团队逐渐被持续的、同质化、低信息量、欠语义化的告警所击溃，狩猎真实威胁犹如大海捞针。通过自动化的告警分类分级技术，能够极大程度上缓解对攻防专家经验的依赖，提升安全运营中心的投入产出比。

4.3.5.3 思路方案

告警分诊的关键在于充分提取、过滤、组装、推断告警关联的事件上下文，并以可量化、可理解的方式向运营人员提供风险排序值。从上下文自动化构建的角度，可划分为以下多个维度：

- 统计上下文，主要是指告警及其关联实体、行为的统计频率、共现频率建模。一个统计建模的经典假设是：从异常检测和大数定理的角度看，高频次告警所蕴含的威胁信息较少。

- 语义上下文，指告警间的触发时序和组合模式，指示了指定的事件规律或用户行为模式，通过主题分析、词嵌入等基于语言模型的建模方法，能够挖掘潜在的语义关联，提升告警的关联分析语义内涵。
- 信息上下文，指相关网络实体的信息流传递过程。通过系统级的数据、实体及行为标注，结合先验规则和基于图的标签传播算法，以估计、推断敏感数据的关键传播路径。
- 意图上下文，指告警涉及技术的高层战术意图抽象。通过 Kill Chain、ATT&CK 等威胁建模方式，可以把告警直接对应到指定的战术阶段当中。更动态的，可通过抽象的行为模板或统计方法，自动抽取实时数据的抽象意图。

上下文的提取不限于以上方式，关键是从风险驱动的各个维度，包括资产、脆弱性、威胁等，提取告警关联的“故事细节”。细节的丰富程度，决定了告警分诊的置信度参数。如图 26 所示，通过对系统级溯源数据（Provenance）的细粒度图谱构建，结合进程、文件、网络等实体的频率共现关系，能够识别出关联系统行为最异常的告警，有效降低误报告警的影响 [31]。

4.3.5.4 关键挑战

告警分诊和误报识别，是一项综合的风险评估技术。在有限的资源投入下，根据企业环境的风险偏好，有效的进行威胁线索排序，能够支撑安全运营威胁狩猎任务的展开。然而，该项技术仍然面临诸多方面的技术挑战：

风险的量化与目标对齐

如前所述，风险是驱动安全运营相关活动开展的关键指标，告警与事件的调查顺序，亦需要可量化的风险度量模式。从资产、脆弱性、威胁等维度吸纳的数据及其之上的数据模式，如果没有理论或规范驱动的计算度量方案，将导致实际事件处理效果与运营目标的偏离，如 MTTD、MTTR 等指标的恶化。

分诊结果的可解释性

该项技术同样面临大部分数据驱动运营技术的共性问题——可解释性的问题。相对传统依赖黑白名单、预设规则等方式的静态分诊与误报识别方式，数据驱动通过语言模型、图模型等手段，以抽象特征驱动分诊流程。分诊结果需要被一线处置的运营人员所理解，才能进一步支持告警的调查进程。因此，增强分诊数据诊断模型及流程的透明性尤为关键。

4.3.6 态势感知与预警

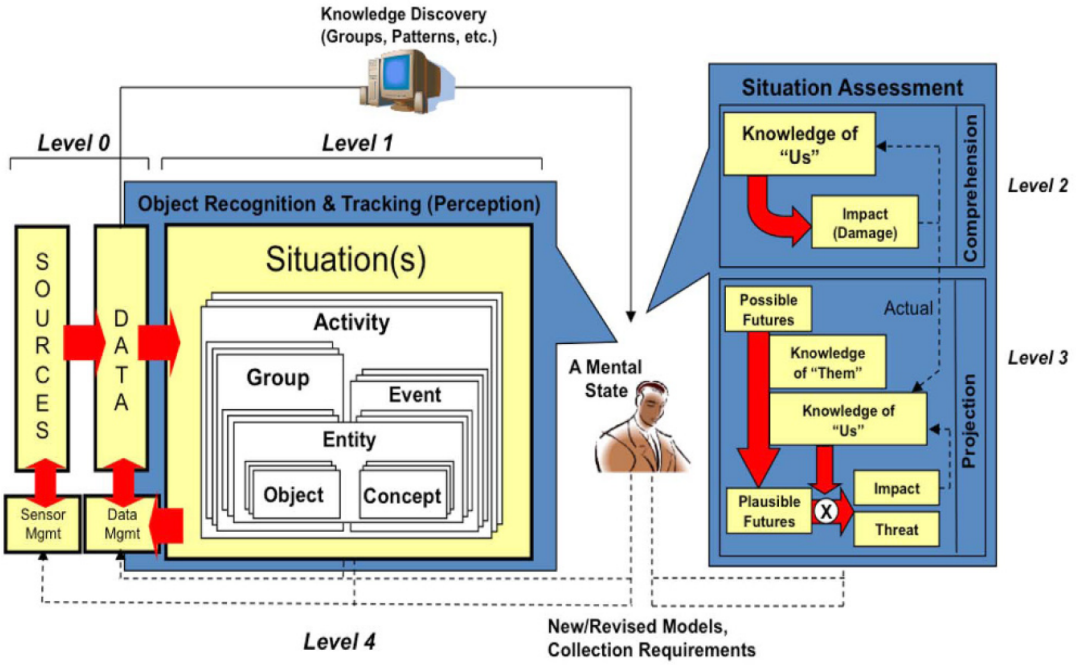


图 27 态势感知参考模型 [32]

4.3.6.1 定义内涵

态势感知与预警的含义是以系统的、整体的、全局的视角，基于网络运行状态数据、情报数据等，抽取、聚合、抽象网络空间关键要素，针对环境变化、攻击意图、行为趋势进行理解，持续监测安全状态，预警可能发生的风险事件，为事件应急处置提供必要的决策依据。

4.3.6.2 技术背景

安全运营的场景已经从点对点的攻击防护取证，拓展到企业组织层次甚至行业级、城市级、国家级平台的安全风险日常分析与管控。安全运营技术不止于细粒度线索驱动的威胁狩猎，还需关注资产、脆弱性、威胁的全局态势，以持续评估安全主体的系统性风险，为网络空间备战、战时决策提供基础。

4.3.6.3 思路方案

态势感知技术的关键在于态势要素的提取、融合、消歧，及基于要素数据的关系推理。从威胁情报

的角度理解态势要素，可包含攻击模式、战役、防护策略、身份、威胁指标、恶意软件、脆弱性、工具、攻击者等等风险关联要素。以网络中攻击者的行为模式为例，通过安全日志、威胁情报数据提取行为特征，并基于特征集合和特征关系的相似程度定义攻击模式，从而将日志数据抽象成攻击行为事件，实现对海量多源异构日志数据的融合并范式化为以攻击模式为主的安全事件，为安全事件分析推理奠定数据基础。在推理方面，可基于融合的知识图谱结构，结合图表示学习、社交网络传播、团伙聚类、路径搜索与推理等方法，在本体实例化数据上完成语义对齐与扩充、攻击链推理、攻击事件聚合溯源等任务，以识别关键局部风险与整体风险点。一个态势感知的流程设计如图 27 所示，该框架基于目标的识别与追踪、态势评估，将态势感知划分能力层级，并构建以知识为核心的感知、理解、融合流程与数据持续迭代的闭环 [32]。

4.3.6.4 关键挑战

态势是一种全局视角，态势的感知与理解是逐层次数据提炼、融合的过程，这一过程中必然丢失细粒度的数据细节。因此态势感知技术需要根据场景需求与目标，在行为、特征可见性与趋势、意图理解任务之间做出取舍。重要的技术挑战包括：

多源态势要素的语义对齐

安全运营的数据源的采集可能执行不同标准，针对同一要素实体的数据抽取因此可具有不同粒度、不同层次、不同标记与命名策略。传统的静态强关联分析方法在大规模数据环境下效果不佳，探索通过语义理解与关键命名实体识别的自动化语义对齐方法尤为关键。

态势与细节的交互性

限于系统处理的瓶颈，态势感知系统一般采用层次化的数据汇聚方式：数据逐层抽象，在中心化管理节点展现全局视图。然而，态势的局部缩放、以及细粒度的事件响应，仍需要提升分布式、层次化数据之间的交互性，保证数据、策略、情报调度的运转效率。

态势风险的量化评估

当前态势评估方法仍然缺乏面向动态风险的标准化、量化的指标体系。要素的状态转变仍以周期性的数据统计为主，亟需数据、情报、知识驱动的推理方法来动态刻画、抽象出可易于被决策者理解的、风险点聚焦的上下文信息，并以高交互的方式呈现。

4.4 鲁棒决策

智能决策的研究一直是人工智能技术研究的圣杯。智能安全运营的决策环节，需要重点关注如何以事件、行为上下文为基础，结合防御知识库，评估事件的综合安全风险，进而给出可行动的执行策略。安全运营决策的关键在于策略生成的鲁棒性。本节将重点介绍风险偏好学习、攻击模拟动态规划和自适应防护策略生成三项关键技术。

4.4.1 风险偏好学习

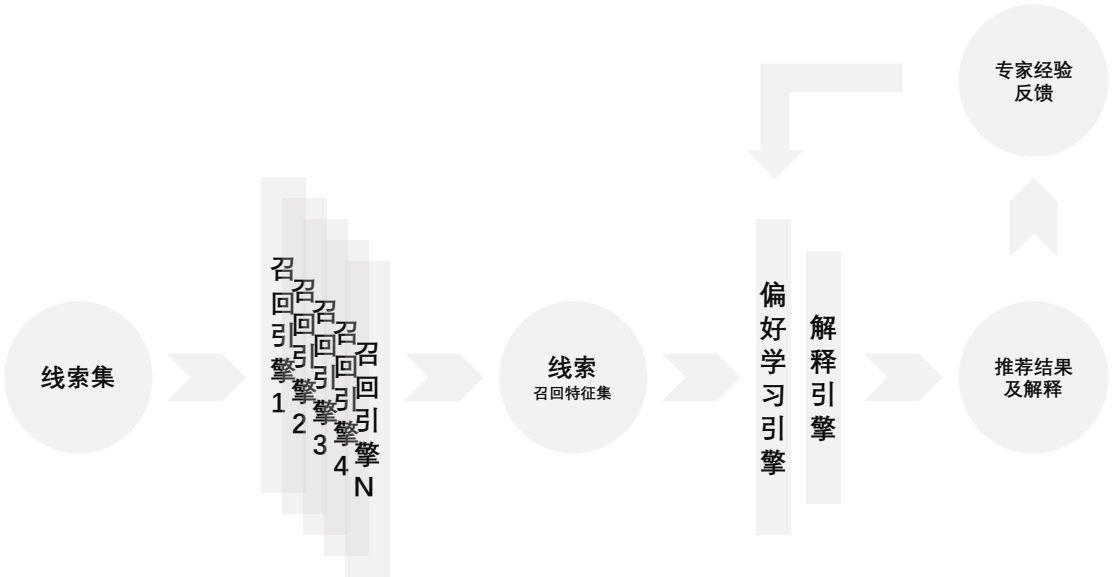


图 28　风险偏好学习引擎架构

4.4.1.1 定义内涵

风险偏好学习的含义是打通人机交互的闭环，通过收集反馈信息，学习专家潜在的、运营导向的风险偏好，识别决定资产、威胁、脆弱性、策略等运营要素风险值的关键数据特征，实现知识先验与数据规律的深度融合，提升系统的决策辅助能力。

4.4.1.2 技术背景

安全运营不同于 IT 系统运维，更强调人与人的对抗。在高度智能化、自动化运营技术成熟之前，安全研究员、运营工程人员的经验与知识发挥着至关重要的作用。然而，现阶段安全数据分析集成平台缺乏交互式设计，系统漏洞和威胁线索的呈现、事件的调查、策略的配置等关键环节难以面向其用户提供个性化的、基于实时风险的运营辅助功能。这导致专家经验难以准确的、快速的转化为系统可识别的机器参数，严重限制了运营技术能力的拓展性和持续新。

4.4.1.3 思路方案

风险偏好学习的关键是面向风险的特征提取与基于用户反馈的偏好拟合。限于时间开销，传统安全运营的驱动力是一些固化的、静态的、基于经验的策略集合。例如特定的漏洞等级、威胁等级与类型等等。而数据层次动态的关联关系、依赖关系，需要通过数据挖掘的方式进行抽取，这些特征通过资产、脆弱性、威胁、防护策略等风险维度进行组织形成风险特征集合，能够向技术平台消费者——运营人员提供数据洞见，辅助事件的理解与策略的选择。进一步，通过构建友好的、可理解的人机交互界面，收集专家在运营流程中的访问行为、偏好分数、页面驻留、描述性反馈等关键信息，在系统后台，基于机器学习或强化学习算法，实现对用户偏好与风险特征集合的数据拟合或自动调整，自适应更新大规模漏洞、资产、线索、事件、策略的动态用户认知风险，最终向运营专家提供量化风险的排序结果。如图 28 所示，是一个典型的偏好学习计算流程。整体上，划分为风险特征召引擎、偏好学习引擎、解释引擎，并通过前端界面的展示与信息采集，形成人机协同的反馈闭环。

4.4.1.4 关键挑战

偏好学习本质上是专家动态标签与网络实体及行为关联特征数据的拟合问题，其基础是数据风险特征集合，而基于特征集的偏好学习过程可通过各类机器学习拟合技术实现。偏好学习模型的优劣决定了决策辅助信息的运营支持效果。技术的关键挑战包括：

交互界面的设计

高质量的、具有明确风险导向性的、机器可读的运营专家反馈信息是偏好学习模型的数据基础。这些数据的积累依赖于系统的技术平台交互界面与数据采集机制的设计。类似有效性评分、偏好选择等显示反馈，以及事件调查页面驻留时间、页面跳转概率等隐性反馈，都是潜在的运营反馈素材，能够在数据拟合阶段发挥标签的作用。

偏好学习的可解释性

风险偏好是一个主观地、较难量化的概念，而复杂的、高维数据拟合可加剧偏好学习的黑盒特性。然而，构建人机协同闭环的关键，决定了偏好学习的输出须具备人可理解的概念解释，这对风险特征集每一维度的可解释性，以及偏好学习模型的可解释性提出了更高的要求。

偏好过拟合与马太效应

鉴于专家反馈标签的语义、规模、时效局限性，偏好学习模型易陷入特征过拟合，导致泛化性能衰减。此外，偏好学习结果与专家反馈的互动也可遭遇马太效应——少量具有特定风险特征的候选类型，如特定类型、指定关联关系的相似告警序列被重复推送，最终导致系统整体的有效覆盖率过低。

4.4.2 攻击模拟动态规划

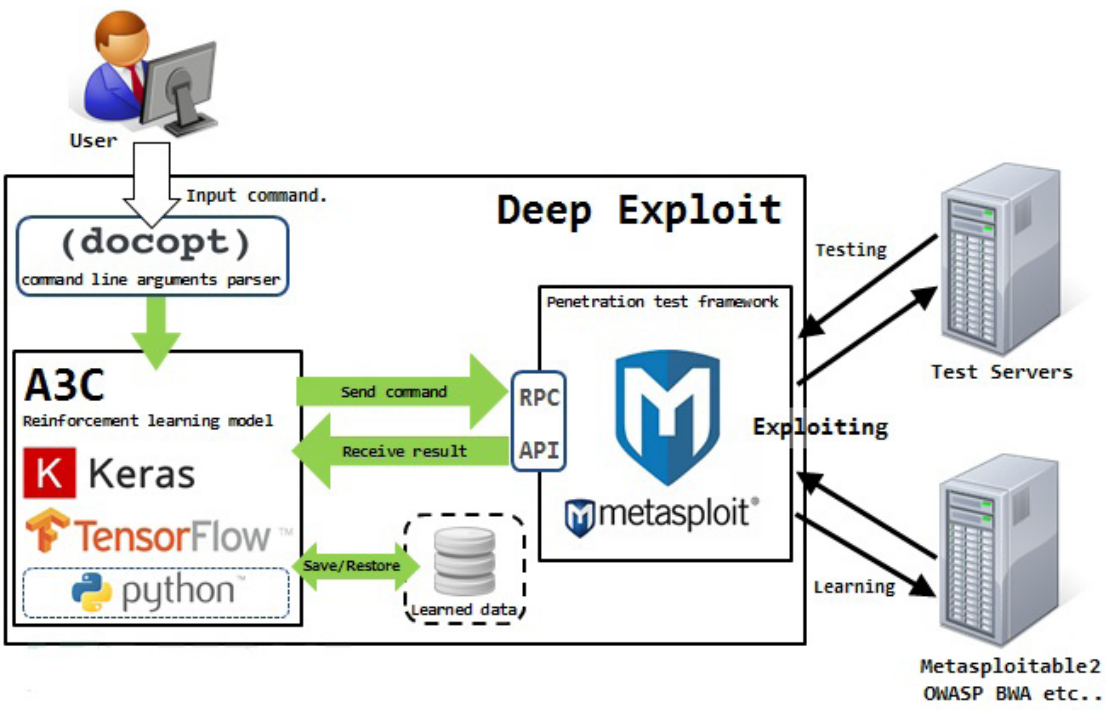


图 29　Deep Exploit 自动化渗透测试 [33]

4.4.2.1 定义内涵

攻击模拟动态规划的含义是基于环境信息和攻击能力图谱，自适应评估攻击模拟效果，实时调整下一步攻击策略、技术实现与路径选择，支撑自动化突破与攻击模拟技术，提升渗透测试、脆弱性评估等主动风险感知运营环节的自动化水平。

4.4.2.2 技术背景

入侵和攻击模拟（Breach and Attack Simulation, BAS）技术成为主动风险感知技术的热点方向。传统渗透测试、漏洞扫描要么过于依赖渗透测试经验，要么受限于有限的环境感知深度。BAS 技术试图通过自动化的渗透攻击脚本，实现机器自主性更高的业务无感知侵入和数据窃取。经典的渗透测试软件，例如 Metasploit，以及 ATT&CK 驱动下的攻击模拟框架，如 Caldera，Infection Monkey 等，已经开始逐渐实现或集成智能化、自动化决策模块，来配合渗透测试工程师实现深入的、环境自适应的高效风险路径、风险数据判断。

4.4.2.3 思路方案

攻击模拟动态规划的关键在于动作、状态、环境以及反馈结果的动态建模，以及基于模型空间的学习过程。在建模方面，核心是规划关联元素的量化表达、交互流程、状态更新函数的设计。如图 29 所示，Deep Exploit 是一个攻击模拟规划技术原型实现 [33]，基于异步优势动作评价（Asynchronous Advantage Actor-Critic, A3C）算法框架，在靶机环境中利用 Metasploit 进行自学习。Deep Exploit 将漏洞利用结果作为奖励函数判断依据，通过大规模的组合测试，使神经网络习得靶标服务器的环境参数与攻击载荷内容之间的潜在映射关系。在学习方法上，动态规划、博弈建模、强化学习、递归贝叶斯估计等经典动态决策框架和算法能够捕获攻击策略选择、多元环境信息与指定攻陷目标函数之间的潜在模式，实现长周期、多阶段的路径自动化规划。

4.4.2.4 关键挑战

相对于攻防对抗实战，攻击模拟演练以环境系统的风险发现为核心目标，面向环境具有相对静态性、可控性，以充分增加入侵攻击的覆盖面与渗透深度。尽管如此，操作系统、服务资源、网络配置、终端防护策略等多维度的网络动态特征，决定了有效的攻击模拟动态规划仍然面临以下挑战：

样本数据自动构建

攻击模拟的自动化过程是动态的博弈过程，所采用的建模方法，例如强化学习，高度依赖训练数据

规模，被动的数据收集模式难以满足复杂模型系统的训练需求。类似棋类、游戏对战博弈智能体训练，需要探索通过在可控状态空间的靶场中构建攻防流程与评估机制，自动化批量生成可供训练输入的样本集合。

攻击策略的泛化性能

专家参与的渗透测试过程，能够根据特定 Web 页面内容、特定系统服务功能，指定针对性的测试载荷，并投递到特定的接口中。自动化攻击模拟限于训练样本空间的有限性，难以有效识别特定的投递入口，并进而生成能够满足业务语义的可用载荷，将大幅限制路径挖掘的深度与广度。攻击策略的泛化需要规划引擎构建鲁棒的可利用单元和功能语义识别能力，同时实时生成可被业务语义成功解析的载荷内容。

4.4.3 自适应防护策略生成

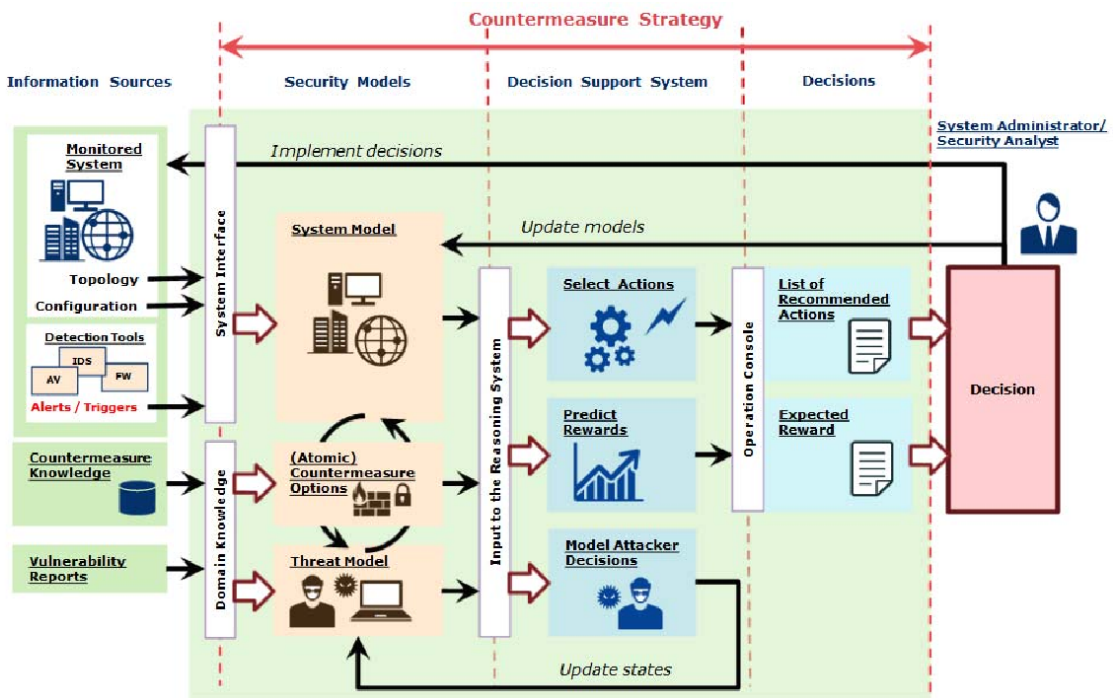


图 30 防护策略推荐框架 [34]

4.4.3.1 定义内涵

自适应防护策略生成的含义是针对持续的线索发现、事件重构、情报命中、脆弱性和资产识别的结果，

基于指定的风险管控目标，动态的从可行防护策略候选列表中选择最佳防护手段，并生成具有可执行参数、步骤、任务依赖的防护策略集合，供运营人员判定或交由调度单元直接下发到指定执行单元。

4.4.3.2 技术背景

SOAR 技术与平台能够快速固化安全运营中的检测与响应知识剧本。但自动化、环境自适应的防御策略选择与生成，面对的是具有高度动态性和对抗性的开放问题和样本空间，静态剧本中策略的实例针对性、有效期、参数适用性等问题在跨场景、跨长时间周期、跨环境的应用中被放大，限制了响应流程的自动化水平。因此，亟需通过数据驱动的、知识驱动、环境驱动方法的融合，结合风险偏好学习机制，制定能够快速有效抑制威胁行为、阻断高风险路径、修复系统损伤，并且不影响正常业务系统运行的实例化防护设备指令。

4.4.3.3 思路方案

一个典型的防护策略生成框架如图 30 所示。自适应防护策略生成的核心在于博弈驱动的策略效果预估与在线策略要素提取。策略效果预估可类比强化学习中的回报函数设计。策略回报的计算需要考虑具体的运营场景。日志或漏洞分诊场景中，漏洞潜在风险、事件规模对人力资源的要求、平均关键任务调查处置时间等因素值得关注；攻击事件响应场景下，对正常业务的误杀率、攻击事件的阻断率、策略执行周期、策略回收周期等因素影响回报的计算结果。核心回报激励计算之外，环境、行动、策略状态空间的构建，也是强化学习等马尔科夫决策框架的重点。防护策略的制定不止于选定特定的策略类型，还需相应的配置策略参数，包括策略自身的阈值、选项、作用域等，以及作用对象的特征、状态、趋势等等。这些策略参数一方面需要结合前述学习过程习得统计性、关联性映射，另一方面需要自适应的数据模式抽取算法，提供在线的、实时的元素特征，技术实现可参考“情报要素的自动化提取”技术章节。

4.4.3.4 关键挑战

策略生成的自动化是安全运营智能化技术体系中最综合的能力体现。一方面该技术的有效实现依赖于精确的线索发现、完整的事件溯源重构、风险偏好的融合等前置环节；另一方面技术的核心实现：动态环境博弈建模与策略学习，是人工智能领域的技术圣杯之一，尚未有成熟的解决方案。

样本空间的局限性

不同于限定策略搜索空间、状态空间下的博弈模型，运营对抗环境下缺乏自动化的攻防样本生成方案，无法批量生成可供强化学习建模的样本集。通过周期性红蓝对抗、靶场攻防模拟可以获得一定数量

的训练样本，但有限的环境配置、攻击手法、策略覆盖等，导致模型面对未知样本时的策略选择偏差。

策略学习的鲁棒性

数据驱动的策略学习过程，需要考虑数据的安全性与对抗安全性。在智能模型攻防研究快速迭代演进的背景下，安全攻防环境的模型鲁棒性尤为关键。攻击者可通过试探性攻击和对抗样本，完成攻击策略层次逃逸（区别于检测逃逸），或造成策略引擎的拒绝服务攻击。因此需要在策略学习的过程中，充分考虑潜在的对抗安全风险。

策略模型的迁移性

策略是场景相关的、平台相关的，并且策略执行的效果“增量”在不同的网络环境下表现不同，例如相同的流量策略执行，不同服务站点的正常业务干扰程度不同。在模型与部署环境的相对迁移过程中，需要充分保留模型核心知识的同时，根据环境和目标需求动态调整策略参数，这对模型自身的可移植性带来挑战。

4.5 可靠行动

智能体的行动不止于一体化的接口设计、集成、与编排。为了保证安全运营行动的持续可靠性，亟需进一步提升策略执行的透明度、可审计性。本节重点介绍透明可审计响应的技术方案的需求与实现。

4.5.1 透明可审计响应

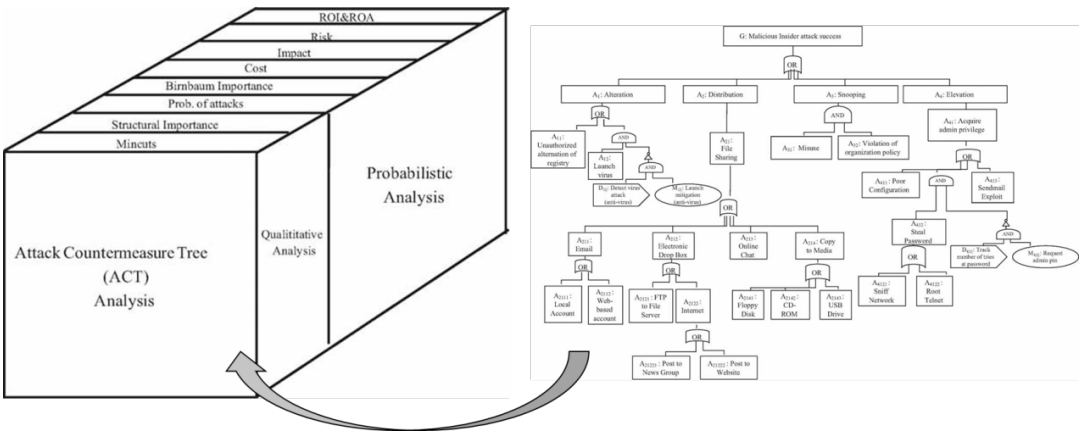


图 31 攻击策略树框架 [35]

4.5.1.1 定义内涵

透明可审计响应的含义是自动化的事件响应需要保持足够的透明度，并提供可供审计的接口与响应审计范本，以在保证系统行动自主性的同时，向运营人员提供完整的、细粒度、结构化、可量化的响应流程、关键数据及其效果反馈，实现自动化响应技术整体可管控。透明可审计响应能力的实现横跨整个智能数据驱动的，是感知 - 认知 - 决策 - 行动的融合体现。按照技术触发的流程环节，归并入负责行动一节。

4.5.1.2 技术背景

智能安全运营的自动化流程的有效运转，依赖多层次技术的鲁棒集成，更离不开人对技术自动化运行过程的监控、反馈、控制与审计。网络空间实体元素的稳定安全，是经济、军事、安全攸关的重要组成。在发挥安全机器智能在特定运行环境下的自主性的同时，唯有透明的、可控的系统行动，才能有效与运营整体流程进行整合。因此，提升事件响应等策略执行环节的信息透明度，是运营技术能力升级的关键环节。

4.5.1.3 思路方案

行动响应透明可审计的关键在于关联技术的透明可解释性、行动目标一致性判定及结构化响应报告生成。行动响应（告警分诊、事件响应、故障恢复）的执行依赖多个前置技术能力，这些技术能力的实现过程中需要兼顾模型、方法的可解释性，具体可参考前述章节，不在此赘述。策略的部署执行的效果，需要行动单元驱动感知单元、认知单元和决策单元，共同收集并判定，以有效监控、评估与预期目标的偏差量。最后，在行动阶段，需要持续汇集决策输出、响应状态、环境反馈等维度的响应要素度量值，并通过结构化、指标化形式的响应审计报告。如图 31 所示，防护策略树（Attack Countermeasure Trees , ACT）框架 [35] 通过构建量化的策略决策体系，并以树形结构组织策略的触发条件与依赖关系，能够以精确的、因果导向的方式表达、概述行动流程。处了树模型之外，基于马尔科夫框架的、基于因果依赖图的结构化响应概述方法，都能够有效融合多维度策略响应元素，形成可解释、可审计的响应反馈数据结构。

4.5.1.4 关键挑战

策略部署执行、效果反馈，有复杂的技术依赖。现阶段 SOAR 的编排与自动化响应能力，可称为流程自动化，距离数据、智能驱动的技术自动化还有很长的路要走。透明可审计响应技术的实现，面临的主要挑战包括：

运营数据生命周期跟踪

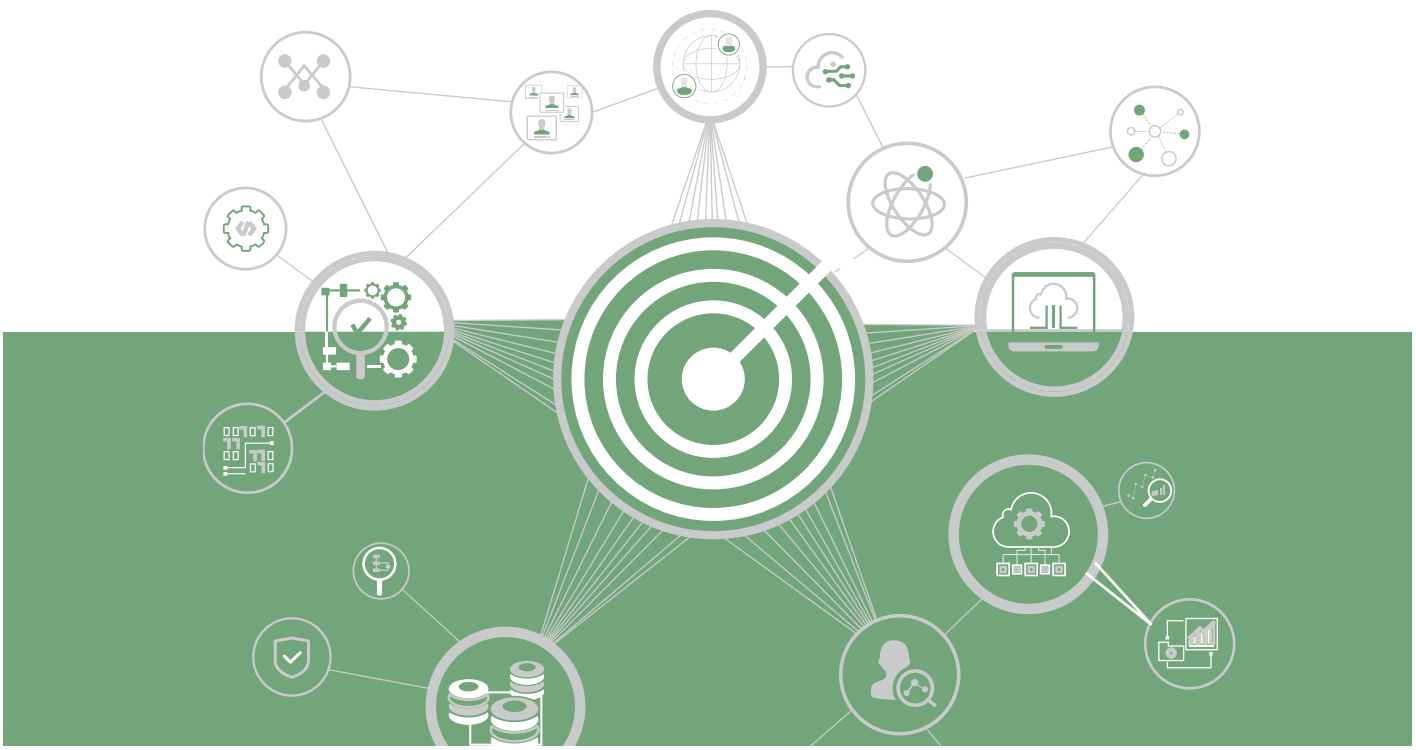
响应行动是自动化系统运行循环中的环节之一，其策略输入、效果输出与整个运营流程在数据层次呈现紧耦合。不确定的网络攻防系统环境中，线索发现与分诊、事件溯源重构、系统风险评估、策略生成等阶段中，运营流程以数字化的形式流转。模型偏差的积累、技术节点的失效等多种因素，都可能导致无效的甚至错误的行动。因此，对运营数据的整个生命周期进行持续监控、分析尤为关键。

策略执行效果度量

网络系统环境的高度动态性与复杂依赖性，导致策略部署之后，对系统产生的实时效果难以形成具有明确因果的指标化度量。除了数据备份、情报富化等静态的处置场景之外，威胁事件阻断、网络隔离、系统热补丁等实时响应执行，可引发关联事件与数据涌现。从复杂、大规模背景数据中识别、召回策略的直接、间接关联数据，是模式识别、因果分析技术领域的难题。

5

AISecOps技术发展趋势



罗马不是一日建成的，AISecOps 技术能力的构建也不能简单的复制其他业务的经验就能一蹴而就。实际上，当今最火热最成熟的人工智能技术，应用场景铺的面非常广，但是应用深度却显不足。类似智能语音服务、图像识别这类服务，是典型的智能化场景，却也只限制于较为低层次的感知层面的任务。在任何自动化过程中需要关键任务决策的，安全、经济、政治、甚至生命攸关的技术场景，如军事、金融、医疗、自动驾驶、法律判决等等，当前的人工智能技术仍难以有效胜任，只能应对场景中的部分问题，距离高度的任务自动化相去甚远。网络安全运营正是此类场景之一。可以说，当前智能化技术本身的不成熟，难以赢得人的信任，成为限制其在许多场景下深入应用的关键问题。本文将分别从可信安全智能技术体系和 AISecOps 技术生态的构建两个方面，展望安全运营技术智能化发展的未来。

5.1 构建可信任安全智能技术体系

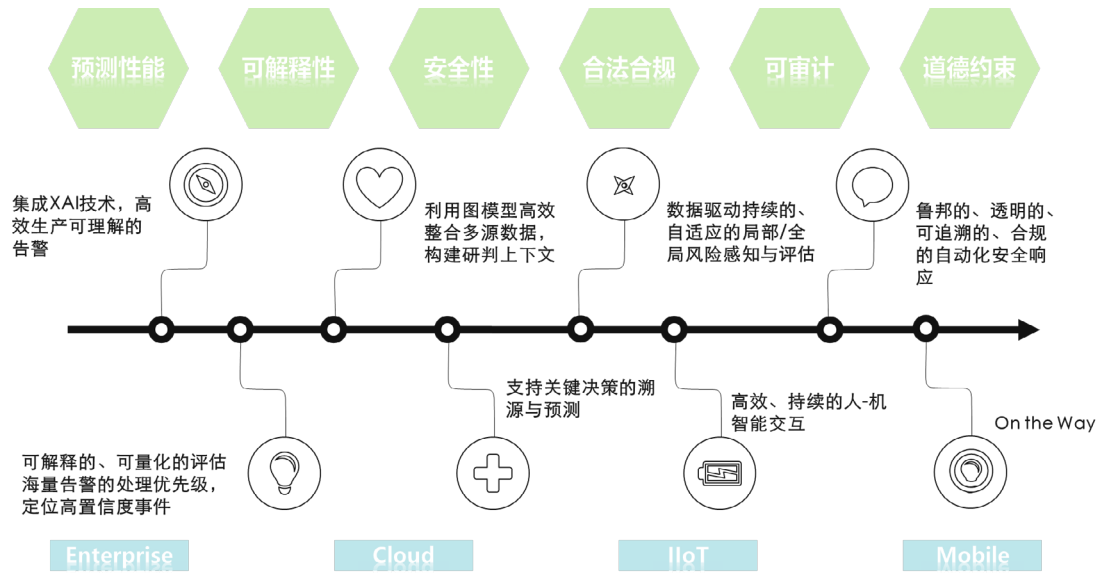


图 32　可信任安全智能体系的技术要素

无论如何，打造更可信任的人工智能，弥补人在处理海量数据过程中的先天不足，打造可信的智能“战友”，始终都是我们的终极追求。人工智能技术在网络安全中的应用，一方面，可以“直接拿来”，应用到网络安全数据分析的非核心场景和流程上，辅助安全工作，如使用自然语言处理技术分析威胁情报或者构建专家系统的对话机器人；使用成熟的图像处理技术检测恶意图片、视频等等；另一方面，需要“优化打造”，构建针对威胁检测、评估、关联、响应等阶段的核心安全智能。如图 32 所示，从构建

技术信任的角度，以提升关键安全能力自动化水平为目标，可信任的安全智能体须具备满足以下核心技术要素的要求，包括预测性能能够适应高度动态的网络环境和攻击场景，模型算法需透明可解释、鲁棒安全、保护隐私，智能技术的执行过程和结果需合法合规、可审计，并在决策执行中满足社会道德约束。以上多个技术要素，互为补充又相互依赖，需要在设计之初充分考虑。正如我们更倾向选择能力强、善于沟通、抗压能力强、高尚守法的人作战友，具备以上技术要素的机器智能更能够获取人的信任，并胜任高级别的安全运营自动化任务。

在可信任安全智能体系的探索过程中，需要充分融合可解释人工智能 XAI、隐私保护技术、图挖掘与分析、智能决策系统、风险评估、人机交互等多学科、多领域智能化技术能力，为安全运营的感知、认知、决策和行动多阶段的任务赋能。

5.2 共建 AISecOps 技术生态

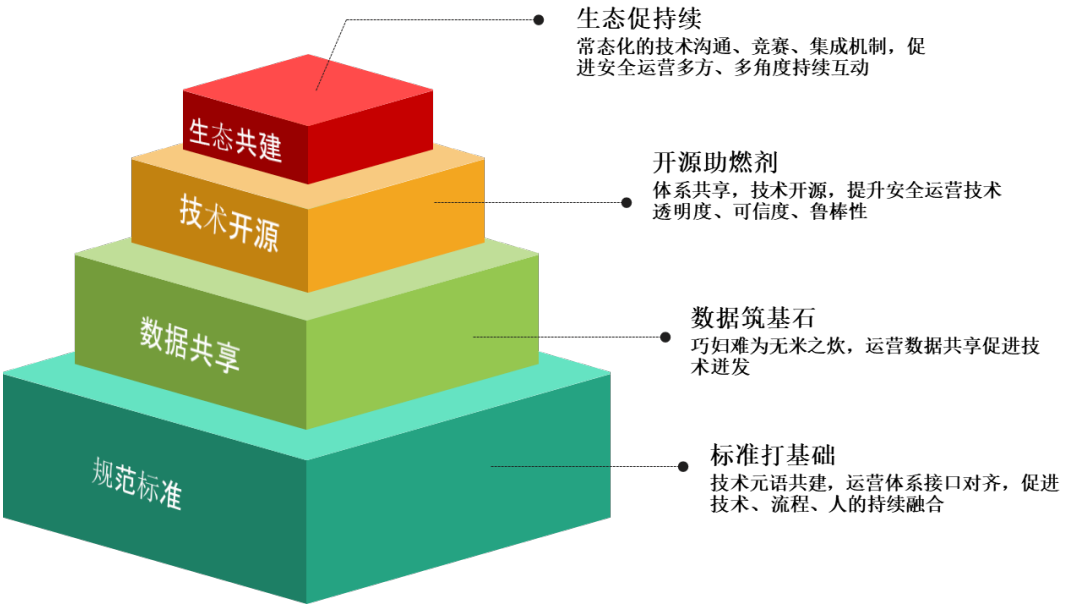


图 33　AISecOps 技术生态的构建

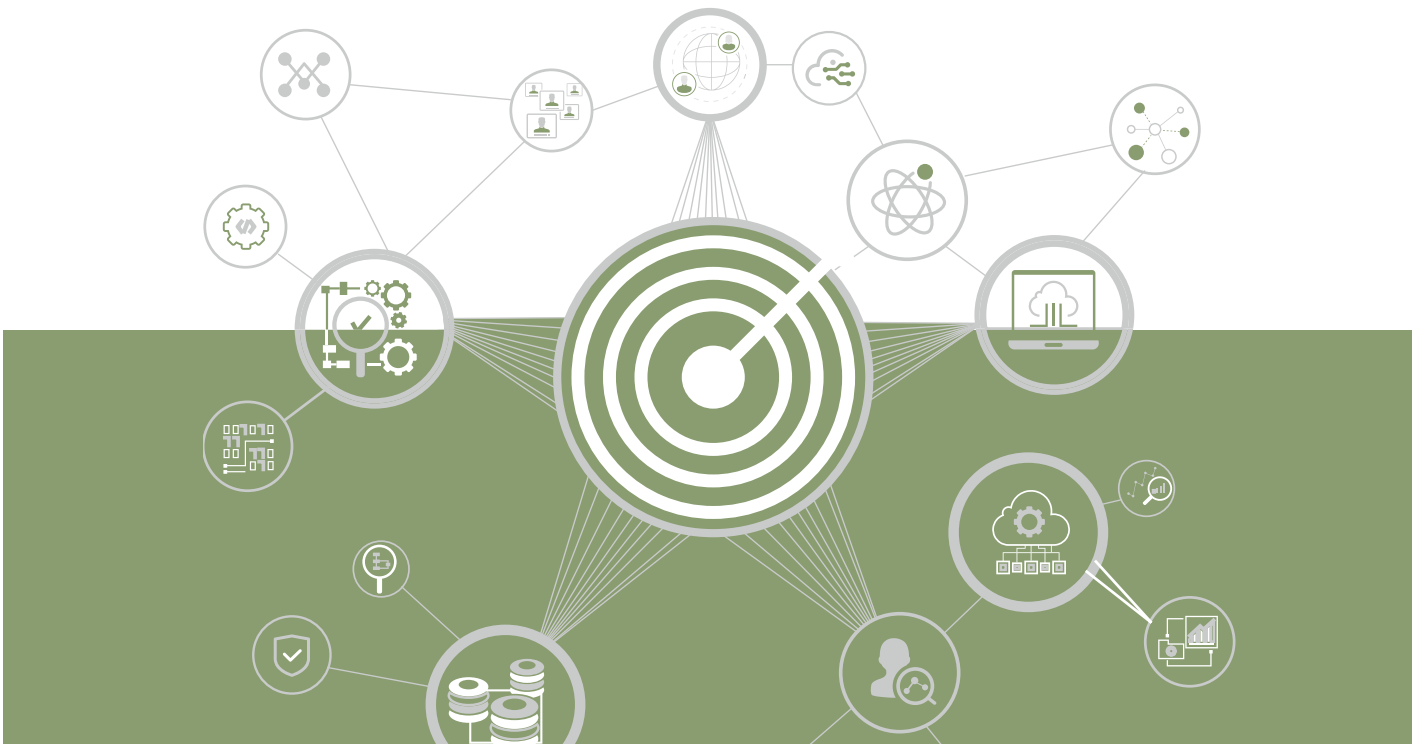
网络安全关系到国计民生，国际形势在快速变化，潜在攻击组织可能通过任何一个脆弱点打破网络安全防线。因此，安全运营看似是企业、组织的自家事，实际上是关系到国家安全的关键环节。从近年来国家级网络威胁情报体系的建设、城市级安全运营中心的雨后春笋般的落地，可窥网络安全运营建设

的国家战略眼光。随着安全运营体系的扩展，所需防护资产从个人、企业、组织的私有资产，逐渐扩大到地区级、国家级关键基础设施及数据资产。此外，技术的发展需要技术生态的构建，以提供持续的、全方位的资源支撑，AISecOps 技术的发展也不例外。现阶段，安全厂商林立，安全技术方案丛出不穷，然而在数据、技术、平台等多层上缺乏规范性约束，在技术交流上缺乏共识元语，在共享合作商缺乏监管机制，多个方面的统一生态的缺失，造成了技术发展的减速。因此，无论从现实需求还是从技术发展的关键途径来看，AISecOps 智能安全运营技术生态体系的建立已是迫在眉睫。

如图 33 所示，本文将概括为规范标准、数据共享、技术开源、生态共建这四个层次。通过行业级、国家级标准的建设，对齐、统一、规范安全运营的关键流程与技术接口，以有效在统一体系下明确分工，优中取优；通过数据共享，建立安全运营技术的标准试验场，促进技术能力的比武与竞争发展；通过技术开源，带动关联技术社区的繁荣，吸引和培养更多安全运营技术人才；最后，通过交流平台与机制共建，打通沟通壁垒、降低合作门槛，真正实现常态化、持久化、全方位的技术交流。

6

总结

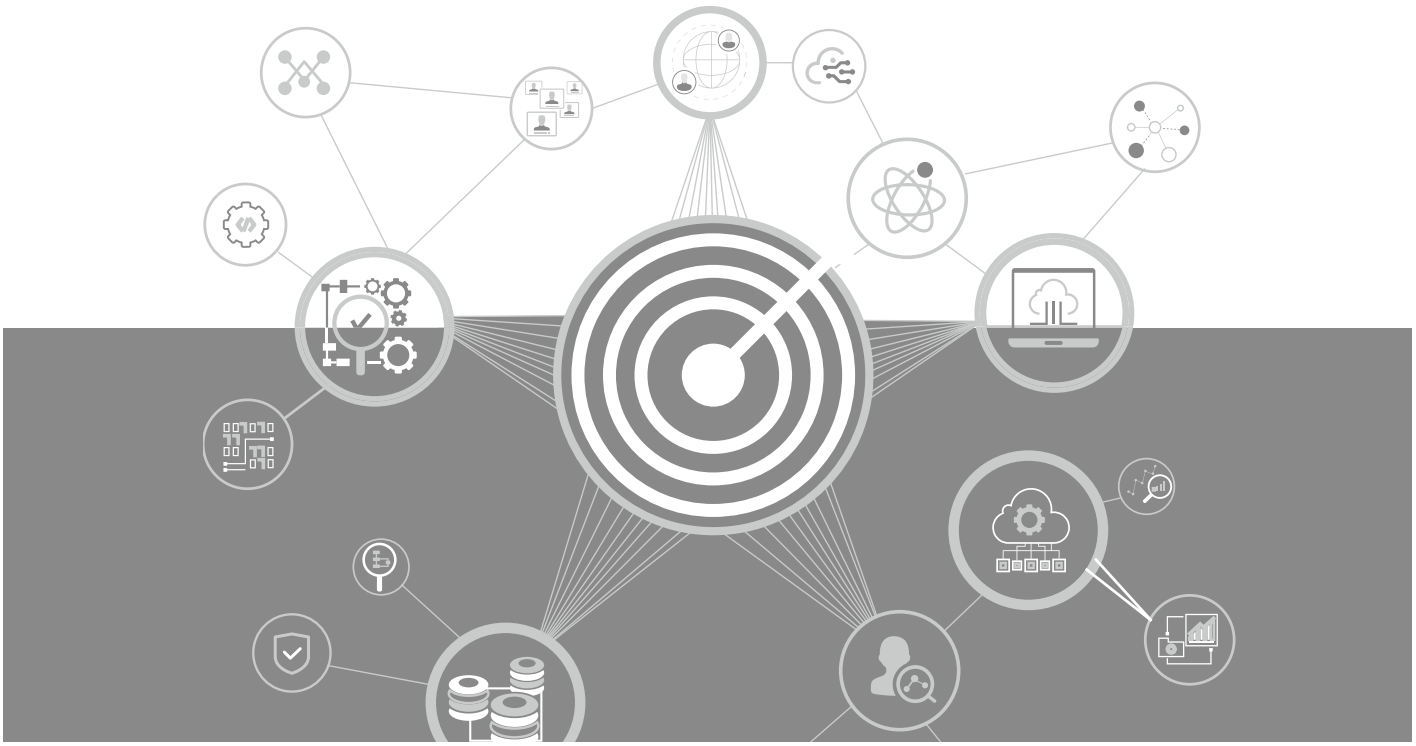


总结

网络安全技术发展已进入以安全风险全生命周期自适应管控与运营为核心的新阶段，面对大规模、多源、高维运营数据的涌入与融合，构建可信任的、可运营的智能安全运营技术体系，支撑网络安全防御体系迈向高度智能化、自动化，解放安全运营的生产力，已成为新基建数字安全时代的重要技术课题。白皮书全面分析了网络安全运营大数据挖掘所面临的关键技术挑战，提出 AISecOps 智能安全运营技术体系。从安全运营的实践出发，深度总结 AISecOps 技术内涵、指标体系、成熟度矩阵、数据分类、技术架构，系统性总结十六大关键基础性技术，并展望了 AISecOps 技术未来发展趋势。期望白皮书能够促进 AISecOps 技术体系的成熟与行业生态的共建，为网络安全运营技术的发展提供实践驱动的基础推动力。

7

参考文献



参考文献

[1] Security Operations Primer for 2020, Gartner, https://www.gartner.com/en/documents/3978969/security-operations-primer-for-2020

[2] Li Z, Chen Q, Yang R, et al. Threat Detection and Investigation with System-level Provenance Graphs: A Survey[M]. arXiv preprint arXiv:2006.01722, 2020.

[3] Noel S, Harley E, Tam K H, et al.: CyGraph: graph-based analytics and visualization for cybersecurity, Handbook of Statistics: Elsevier, 2016: 117-167.

[4] Hassan W U, Bates A, Marino D. Tactical Provenance Analysis for Endpoint Detection and Response Systems[C]. 2020 IEEE Symposium on Security and Privacy (SP), 2020: 1172-1189.

[5] Shen Y, Stringhini G. ATTACK2VEC: Leveraging Temporal Word Embeddings to Understand the Evolution of Cyberattacks[C]. USENIX Security Symposium, 2019.

[6] Guo W, Mu D, Xu J, et al. LEMNA: Explaining Deep Learning based Security Applications[C]. Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018: 364–379.

[7] Gunning D. Explainable artificial intelligence (xai)[J]. Defense Advanced Research Projects Agency (DARPA), nd Web, 2017, 2(2).

[8] https://www.plugandplaytechcenter.com/resources/machine-learning-security-3-risks-be-aware/

[9] Dang Y, Lin Q, Huang P. AIOps: real-world challenges and research innovations[C]// 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), 2019: 4-5.

[10] Rowley, J. The wisdom hierarchy: representations of the DIKW hierarchy[J]. Journal of information science, 2007, 33(2): 163-180.

[11] https://attack.mitre.org/

[12] https://capec.mitre.org/

[13] https://cwe.mitre.org/

[14] Grant T. Unifying planning and control using an OODA-based architecture[C]. Proceedings of Annual Conference of the South African Institute of Computer Scientists and Information Technologists, 2005: 111-122.

[15] https://en.wikipedia.org/wiki/Self-driving_car

[16] https://oasis-open.github.io/cti-documentation/

[17] Jajodia S, Noel S, Kalapa P, et al. Cauldron mission-centric cyber situational awareness with defense in depth[C]. MILCOM 2011 Military Communications Conference, 2011. 1339-1344.

[18] Shu X, Araujo F, Schales D L, et al. Threat Intelligence Computing[C]. Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018: 1883-1898.

[19] 基于知识图谱的 APT 组织追踪治理，绿盟科技，https://cloud.tencent.com/developer/article/1556638

[20] Enhanced Cyber Threat Model for Financial Services Sector (FSS) Institutions

[21] https://github.com/mitre/caldera

[22] https://github.com/endgameinc/RTA

[23] Runzi Zhang, Mingkai Tong, Lei Chen, et al. CMIRGen: Automatic Signature Generation Algorithm for Malicious Network Traffic. TrustCom 2020.

[24] 郭莉 , 曹亚男 , 苏马婧 , 等 . 网络空间资源测绘 : 概念与技术 [J]. 信息安全学报 , 2018, 003(004):1-14.

[25] Liu T , Qi Y , Shi L , et al. Locate-Then-Detect: Real-time Web Attack Detection via Attention-based Deep Neural Networks[C]. Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19. 2019.

[26] Wang Q, Hassan W U, Li D, et al. You are what you do: Hunting stealthy malware via data provenance analysis[C]. Symposium on Network and Distributed System Security (NDSS), 2020.

[27] https://www.silverfort.com/blog/detecting-and-predicting-malicious-access-in-enterprise-networks-using-the-louvain-community-detection-algorithm

[28] Milajerdi S M, Gjomemo R, Eshete B, et al. HOLMES: real-time APT detection through correlation of suspicious information flows[J]. arXiv preprint arXiv:1810.01594, 2018.

[29] Hossain M N, Sheikhi S, Sekar R. Combating Dependence Explosion in Forensic Analysis Using Alternative Tag Propagation Semantics[J].

[30] Pitropakis N , Panaousis E , Giannakoulias A , et al. An Enhanced Cyber Attack Attribution Framework[J]. 2018.

[31] 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.

[32] Jajodia S , Liu P , Swarup V , et al. Cyber Situational Awareness: Issues and Research[J]. 2009.

[33] https://github.com/13o-bbr-bbq/machine_learning_security/tree/master/DeepExploit

[34] Nespoli P , Papamartzivanos D , Marmol F G , et al. Optimal countermeasures selection against cyber attacks: A comprehensive survey on reaction frameworks[J]. IEEE Communications Surveys & Tutorials, 2018:1-1.

[35] Roy A , Kim D S , Trivedi K S . Attack countermeasure trees (ACT): towards unifying the constructs of attack and defense trees[J]. Security and Communication Networks, 2012, 5(8):929-943.

绿盟科技创新中心

绿盟科技创新中心是绿盟科技的前沿技术研究部门。包括云安全实验室、数据分析实验室和物联网安全实验室，关注云安全、容器安全、威胁情报、数据驱动安全、物联网安全和区块链等领域。作为“中关村科技园区海淀园博士后工作站分站”的重要培养单位之一，与清华大学进行博士后联合培养，科研成果已涵盖各类国家课题项目、国家专利、国家标准、高水平学术论文、出版专业书籍等。我们持续探索信息安全领域的前沿学术方向，从实践出发，结合公司资源和先进技术，实现概念级的原型系统，进而交付产品线孵化产品并创造巨大的经济价值。

天枢实验室

天枢实验室聚焦安全数据、AI 攻防等方面研究，以期在“数据智能”领域获得突破。

绿盟科技威胁情报中心

绿盟科技威胁情报中心（NSFOCUS Threat Intelligence center, NTI）是绿盟科技为落实智慧安全 2.0 战略，促进网络空间安全生态建设和威胁情报应用，增强客户攻防对抗能力而组建的专业性安全研究组织。其依托公司专业的安全团队和强大的安全研究能力，对全球网络安全威胁和态势进行持续观察和分析，以威胁情报的生产、运营、应用等能力及关键技术作为核心研究内容，推出了绿盟科技威胁情报平台以及一系列集成威胁情报的新一代安全产品，为用户提供可操作的情报数据、专业的情报服务和高效的威胁防护能力，帮助用户更好地了解和应对各类网络威胁。

NSFOCUS

THE EXPERT
BEHIND GIANTS
巨人背后的专家

二十年来，绿盟科技致力于安全攻防的研究，
为政府、运营商、金融、能源、互联网以及教育、医疗等行业用户，提供
具有核心竞争力的安全产品及解决安宁，帮助客户实现业务的安全顺畅运行。
在这些巨人的背后，他们是备受信赖的专家。

www.nsfocus.com

欢迎关注
绿盟科技官方微信