

基于告警增量分析的攻击场景实时重构方法

赵新建¹, 汤慧敏², 陈石¹, 张玉健², 张颂¹, 程光²

(1. 国网江苏省电力有限公司信息通信分公司, 江苏 南京 210024;

2. 东南大学网络空间安全学院, 江苏 南京 211189)

摘要: 面对日益复杂的多阶段、长周期网络攻击, 安全事件常表现出跨时空域的交织关联。攻击场景重构能够有效支撑溯源取证和威胁发现, 然而现有在攻击结束后基于全量告警数据分析的方法, 存在结果滞后和效率低等问题。为此, 针对告警信息的增量更新和全局视角存在矛盾问题, 提出了基于时间窗口的增量式攻击场景重构方法。该方法的主要技术创新如下: ①通过预处理和时间窗口划分, 实现告警数据的规范化表达和分析规模的有效控制; ②针对单窗口增量更新数据, 以目标资产为中心构建攻击告警图并挖掘频繁1-项集, 在此基础上通过图遍历策略生成并筛选所有可能攻击路径; ③针对多个窗口的全局视角数据, 提出一种攻击路径融合算法对已有路径和增量路径进行关联, 实现攻击场景的全局重构。在开源和真实数据集上的实验结果表明, 所提方法能够有效实现攻击场景重构, 同时相比全量分析方法显著降低了计算复杂度, 在增量场景下保持了稳定的处理性能。

关键词: 攻击场景重构; 多步攻击; 告警关联; 增量更新; 时间窗口

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.AQ25274

Real-time attack scenario reconstruction method based on incremental alert analysis

Zhao Xinjian¹, Tang Huimin², Chen Shi¹, Zhang Yujian², Zhang Song¹, Cheng Guang²

1. Information & Telecommunication Branch, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China

2. School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China

Abstract: In the face of increasingly complex multi-stage and long-term cyber attacks, security events often exhibit interconnections across both temporal and spatial domains. Attack scenario reconstruction plays a crucial role in supporting traceability, forensics, and threat discovery. However, existing methods that rely on analyzing full alert data after an attack has concluded suffer from issues such as delayed results and efficiency bottlenecks. To address the inherent tension between incremental alert updates and a global perspective, an incremental attack scenario reconstruction method based on time window was proposed. The key technical innovations of the proposed method were as follows: ① Through preprocessing and time-window partitioning, normalized representation of alert data was achieved while effectively controlling the analysis scale; ② For incrementally updated data within a single window, an attack alert graph was constructed centered on target assets as well as mining frequent 1-itemsets, followed by the use of a graph traversal strategy to generate and filter all possible attack paths; ③ For multi-window data from a global perspective, an attack path fusion algorithm was introduced to correlate historical paths with incremental paths, enabling global reconstruction of the attack scenario. Experimental results on both open-source and

收稿日期: 2025-11-29; 修回日期: 2026-02-14

通信作者: 程光, chengguang@seu.edu.cn

基金项目: 国网江苏省电力有限公司科技项目 (No.J2024086)

Foundation Item: The Science and Technology Project of State Grid JiangSu Electric Power Company (No.J2024086)

real-world datasets demonstrate that the proposed method effectively achieved attack scenario reconstruction while significantly reducing computational complexity compared to full analysis method, maintaining stable processing performance in incremental scenarios.

Key words: attack scenario reconstruction, multi-stage attack, alert correlation, incremental update, time window

0 引言

随着网络攻防技术的不断演进,网络攻击不仅呈现出手段多样化的特征,还具备多阶段、长周期的特性。以高级持续性威胁(advanced persistent threat, APT)攻击^[1]为例,其不仅融合了社交攻击、网络钓鱼及零日漏洞利用等技术手段,还采用了多阶段攻击模式,从前期的信息收集到达成最终目标,整个攻击生命周期往往持续数月甚至更长时间。近年来,此类攻击不断威胁各国的电力、能源等关键基础设施安全,2015年乌克兰大面积停电事件、2019年委内瑞拉电力网络瘫痪事件^[2]均是APT攻击引发的典型安全事故,因此,针对该类型多阶段、长周期攻击的有效检测问题亟待解决。

攻击场景重构作为网络攻击检测与分析的有效手段之一,旨在通过对入侵检测系统(IDS)等安全设备产生的原始安全事件进行关联分析,还原攻击者对整个网络空间的攻击与渗透过程^[3]。其中,告警(Alert)是指网络安全设备基于规则或模型自动化检测所发现的可疑行为(如暴力破解、结构化查询语言(SQL)注入、Web-Shell上传等)记录。由于这些告警信息经过网络安全设备的研判和标准化,因此具有较高的应用价值。在此基础上,通过告警关联分析,可以将同一次攻击活动中多个不同类型的单步攻击产生的告警进行关联,生成以攻击意图划分的攻击阶段,从而实现攻击场景重构。然而,在实际网络环境中,连续的攻击和不相关的攻击可能同时发生,导致存在因果关联的告警与不相关的告警交织在一起;同时,攻击者未必按照精确的时间顺序执行攻击,因此仅依靠采集数据的时间序列分析难以发现真正的因果关系。在面对多阶段和多目标特性的攻击时,精确描述脆弱性组件之间的关联变得尤为重要。

1 相关工作

根据所采用技术路径的不同,现有基于告警

信息的攻击场景重构方法可以分为专家经验、因果关联、相似性分析和机器学习4类^[4],各类方法均具有独特的优势和局限性。

基于专家经验的方法利用领域专家知识和预定义规则来构建攻击模式识别框架。Bryant等^[5]提出一种基于混合杀伤链的攻击模板配置方案,以提高安全信息和事件管理的性能;Barzegar等^[6]提出基于本体的专家系统方法,通过推理告警流中的证据来重构攻击场景;Kenaza等^[7]提出一种基于描述逻辑的本体表示方法,为入侵检测中的事件关联提供知识表示和推理框架。这类方法的优势在于构造精准、关键攻击环节辨识度高,但其应用效果往往依赖于专家知识库所定义的攻击关联规则与专家经验的完整性和准确性,因此人工参与程度较高。

基于因果关联的方法通过构建告警信息间的因果关联模型,逐步还原攻击者的操作步骤。王英州等^[8]提出一种针对工业控制系统物理层告警量关联的算法,利用该算法得到的全局告警关联结构具有较好的可解释性,有助于分析告警量在物理层的关联关系。Yang等^[9]提出一种名为Poirot的因果关联辅助语义分析系统,该系统能够自主挖掘异常事件之间的因果关系,构建告警链;Zang等^[10]融合来自安全信息和事件管理系统的异构威胁情报,通过分析威胁情报的因果关系重建多阶段攻击场景;张玉臣等^[11]利用前后依赖关系和告警的时间序列对识别出的攻击事件进行前后关联;Zhao等^[12]利用告警日志预测潜在的攻击目标,采用强化学习确定告警数据之间的因果关系。这类方法表达直接、交互性强且可拓展性好,但重构效果依赖于因果关系模型设计的合理性。

基于相似性分析的方法通过计算不同告警之间的相似度来定义其关联程度。Wang等^[13]根据告警语义和攻击阶段的相似性聚合了多步攻击路径;Wang等^[14]设计了5类与载荷相关的相似度因子,用于计算告警之间的相似度;Haas等^[15]提出了一种基于图的告警关联算法,该算法基于

主机通信对告警进行聚类 and 连接, 以实现多阶段重构; Hu 等^[16]将告警映射到攻击图并输出攻击序列, 通过合并断裂的攻击场景来优化重构效果。这类方法能够识别已知场景的变体, 但效果依赖于相似性计算方法和已有攻击场景库的完整性, 对未知攻击场景的重构能力有限。

基于机器学习的方法利用数据挖掘和深度学习技术自动挖掘攻击模式和事件关联, 用于替代专家经验和人工规则。Jia 等^[17]采用少样本深度学习算法及双向循环神经网络模型来检测 APT 攻击中的佯攻行为; Farhadi 等^[18]提出一种算法用于从警报流中发现多阶段攻击, 并通过隐马尔可夫模型预测未来事件; 吕磊等^[19]将遗传算法与广义序贯模式算法相结合, 从海量告警信息中挖掘频繁序列模式; Mao 等^[20]通过整合攻击时间、风险评估和攻击节点信息, 实现主机间攻击场景重构。这类方法移植效率高、优化空间大, 但效果依赖于模型结构的设计和训练数据的质量。

现有基于告警信息的攻击场景重构方法虽从不同技术路径实现了攻击场景的有效重构, 但普遍依赖全时段、全量告警数据分析, 需要等待攻击活动完全结束后才能进行完整分析, 且分析过程往往需要遍历全时段告警信息, 计算复杂度随数据量的增长显著提升, 难以满足网络攻击实时研判与应急处置的需求。

针对上述问题, 本文提出一种基于时间窗口的增量式攻击场景重构方法, 即实时攻击场景重构 (real-time attack scenario reconstruction, RASR)。其主要思路如下: 首先, 对原始告警进行预处理并划分时间窗口, 实现告警数据的规范化与处理规模的有效控制; 随后, 在每个独立窗口内, 围绕目标资产构建告警图并挖掘频繁 1-项集, 再通过图遍历枚举所有可能的攻击路径并对其进行筛选; 最后, 设计一种跨窗口的攻击路径融合算法, 将已有路径与增量路径进行关联, 完成攻击场景的全局化重构。

本文的主要贡献如下。

1) 提出一种基于告警图的攻击路径生成方法, 能够准确识别告警间的因果关联, 有效挖掘潜在攻击路径, 为攻击场景重构提供基础。

2) 提出基于窗口划分的告警增量分析方法,

按窗口分段处理告警数据, 并通过增量方式扩展攻击路径, 实现高效、实时的攻击场景重构。

2 攻击场景实时重构模型建立

2.1 攻击场景实时重构威胁模型

威胁模型是基于专家经验和实际业务, 对网络攻击过程进行抽象、描述和分析的形式化表达。其作用是对威胁进行结构化表示, 并为攻击场景重构提供指导。本文参考 MITRE ATT&CK 威胁模型, 该模型的更新频率较高, 其最新版本为 2026 年 4 月发布的 v19^[21]。MITRE ATT&CK 根据真实的攻防观测数据构建战术与技法知识库, 可用于网络攻击的识别与分类、归因分析及风险评估等。

本文参考 MITRE ATT&CK 将网络攻击分为 5 个攻击阶段和 14 种攻击战术。攻击阶段包括信息收集、初始入侵、据点建立、控制通道建立和目标主机破坏。一般而言, 攻击者首先利用自动化扫描工具进行漏洞探测, 对目标进行侦察与信息收集; 在形成初步认知后, 通过网络钓鱼、恶意邮件或木马程序等方式实施初始入侵, 建立初始立足点; 随后借助漏洞实现权限提升或部署有效载荷, 构建命令和控制 (command and control, C&C) 通道; 在获得持续控制能力后, 攻击者在网络内部横向移动, 扩大攻击范围。攻击战术包括侦察、资源开发、初始访问、攻击执行、持久化、权限提升、防御规避、凭证获取、环境发现、横向移动、信息收集、命令与控制、数据渗出和危害影响。上述攻击战术可以根据其达成的目标归属到相应的攻击阶段, 对应关系见表 1。

表 1 攻击阶段和攻击战术对应关系

攻击阶段	攻击战术
信息收集	侦察、资源开发
初始入侵	初始访问、攻击执行、持久化、权限提升
据点建立	防御规避、凭证获取、环境发现、横向移动
控制通道建立	信息收集、命令与控制
目标主机破坏	数据渗出、危害影响

2.2 攻击场景实时重构数据模型

数据模型是为实现攻击场景重构, 将告警、流量、日志、主机衍生物等痕迹信息组织为特定

的逻辑结构和存储格式的模型。本文主要关注告警信息。告警信息由网络安全设备对一条或多条异常日志进行聚合后生成,用于对安全事件进行简要描述。它通常以非结构化或半结构化形式存在,一般包含攻击源、攻击目的、攻击方式、攻击描述等信息。

本文所用告警数据主要来源于电力系统网络安全监测设备,在原始告警数据的基础上提取关键参数,并补充后续分析所需的参数。告警信息包括告警编号 (ID)、攻击类型 (alertType)、时间戳 (timeStamp)、源 IP 地址 (sIP)、目的 IP 地址 (dIP)、目标资产名称 (assetName)、资产所属地域 (region)、攻击威胁程度 (degree)、攻击战术 (tactic)、攻击阶段 (stage)、攻击状态 (attackStatus),具体参数说明见表 2。其中,攻击战术和攻击阶段的取值参考 2.1 节定义的威胁模型。

表 2 告警信息参数说明

参数名称	参数描述
ID	告警信息在数据集中的编号
alertType	告警的攻击类型
timeStamp	告警产生的时间
sIP	触发告警的源 IP 地址
dIP	触发告警的目标 IP 地址
assetName	触发告警的资产名称
region	触发告警的资产所属地域
degree	告警的攻击威胁程度
tactic	告警攻击类型所属的攻击战术
stage	告警攻击类型所属的攻击阶段
attackStatus	告警的攻击状态

2.3 实时重构方法优劣的评价指标

围绕上述威胁模型和数据模型,本文所提方法需要满足以下两个核心目标。

1) 还原度:攻击场景重构是从告警等痕迹信息中还原攻击步骤。可以基于已标注的数据集,通过完整性、健全性等量化指标,评估重构方法对攻击场景的识别和还原的效果。

2) 高效性:攻击场景重构需要处理海量告警信息,特别是在长周期攻击场景下,重构方法的分析效率对于提升威胁发现的实时性具有重要意义。

3 攻击场景实时重构方法

本文提出了基于告警增量分析的攻击场景实时重构方法 RASR,其总体架构如图 1 所示,主要包含以下 3 个部分。

3.1 预处理与窗口划分

预处理主要是对告警数据进行规范化处理,具体步骤如下:首先依据预定义的数据模型将原始告警转换为统一的实体对象,并对其进行整合排序;然后通过比较告警的攻击时间、类型、源 IP 地址、目的 IP 地址、端口及攻击状态等特征,去除重复记录;接着根据 ATT&CK 模型,按攻击类型对告警信息进行分类,并据此将告警映射至对应的攻击阶段。经过上述处理,最终得到格式统一、去冗余且字段完整的规范化告警数据。

窗口划分旨在缩减待处理数据的规模。本文采用时间窗口将连续的告警流划分为多个离散的时间片段,对各窗口内的告警进行独立分析,并对窗口间的分析结果进行关联。图 2 给出了一个告警序列经时间窗口划分的示例。初始告警数据被划分为 3 个时间窗口,各窗口大小设为固定

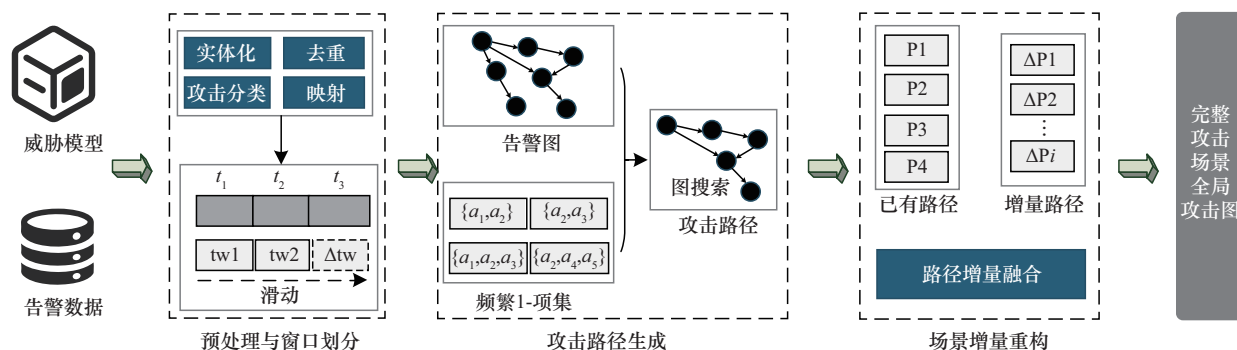


图 1 RASR 总体架构

值，即 $tw = t_2 - t_1 = t_3 - t_2 = t_4 - t_3$ ；后续到达的增量告警数据按相同的间隔 tw （即 $t_5 - t_4 = t_6 - t_5$ ）被划分至新增窗口中进行处理。

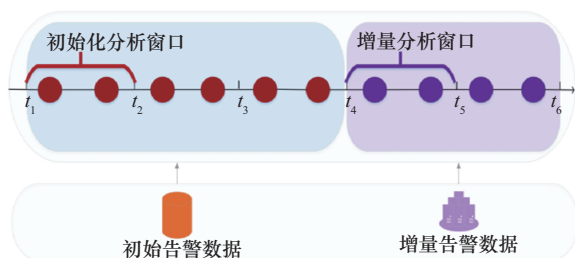


图2 时间窗口划分示例

时间窗口的大小直接影响分析效果，具体体现在计算性能和攻击路径完整性两个方面。理论上，窗口太小易导致攻击路径碎片化，无法捕捉完整的攻击路径，也会增加后续路径合并的开销；窗口太大则会降低告警分析的实时性。本文方法对时间窗口大小无硬性限制。在实际分析场景中，电力物联网安全运维通常采用日报制度，安全分析人员每日汇总前一天的告警数据进行分析，这一行业实践使得 24 h 窗口在攻击路径完整性与分析实时性之间取得良好平衡。因此，本文选取 24 h 作为时间窗口，以贴合实际运维需求。

3.2 攻击路径生成

本节对时间窗口内的告警数据进行分析以生成攻击路径，主要包含告警图构建和频繁 1-项集挖掘两个前序步骤，以及最终的路径生成步骤。各步骤的具体设计如下。

(1) 告警图构建

本文提出了一种基于时序告警关联的多起始点攻击图构建算法，具体如算法 1 所示。对于每个时间窗口内的告警数据，算法以每条告警为节点，目标资产为中心构建告警图。首先，按 (sIP, dIP) 对告警进行分组：

$$G(s,d) = \{ a \in \text{alerts} \mid (a.sIP = s \wedge a.dIP = d) \vee (a.sIP = d \wedge a.dIP = s) \} \quad (1)$$

其中，alerts 为告警集合， a 为该告警集合中的一个告警， $G(s,d)$ 为 IP 对 (s,d) 对应的告警分组。该分组策略将双向通信的告警归为同一组，以支持复杂攻击场景的分析。接着，对每组内的

告警按时间戳排序（行 1~7）。针对每个新到达的告警，算法通过多前驱查找机制识别所有可连接的前驱节点（行 8~18）。该机制按时间倒序遍历所有历史告警（行 11），通过阶段兼容性约束（行 13）筛选有效前驱告警，仅允许阶段前进（支持攻击路径的正常演进）或同阶段持续（支持同一阶段内的并行分支）的连接；同时，算法验证时间顺序性（行 14），确保当前告警时间不早于前驱告警，保证攻击路径的因果关系合理。由于 validPredecessors 可以包含多个前驱节点（行 15），算法能够捕捉攻击路径的“分支-汇聚”结构，支持多个并行的攻击分支汇聚到同一个后续告警，还原复杂的攻击场景。对所有筛选出的有效前驱告警，算法建立有向边连接（行 19~26）；对于无有效前驱的告警（即入度为零），算法自动将其识别为独立的攻击起始点（行 27~30）。最终，算法通过建立适当的“源-目的”节点的连接，生成支持多分支汇聚的有向无环攻击图（行 31~37），完整保留攻击的逻辑关系和时序特征。该算法的时间复杂度为 $O(n^2)$ ，其中 n 为告警数量。

算法 1 多起始点攻击图构建算法

输入 告警集合 alerts

输出 攻击图集合 graphs

```

1: graphs  $\leftarrow \emptyset$ 
2: groups  $\leftarrow \text{GroupByIPPair}(\text{alerts})$ 
3: for each ipPair in groups do
4:   sortedAlerts  $\leftarrow \text{SortByTime}(\text{groups}[\text{ipPair}])$ 
5:   (source, target)  $\leftarrow \text{GetDirection}(\text{sortedAlerts})$ 
6:   graph  $\leftarrow \text{CreateGraph}()$ 
7:   AddAlertNodes(graph, sortedAlerts)
8:   for  $i = 1$  to length(sortedAlerts) - 1 do
9:     current  $\leftarrow \text{sortedAlerts}[i]$ 
10:    validPredecessors  $\leftarrow \emptyset$ 
11:    for each prevAlert in Reverse(sortedAlerts[0:i-1]) do
12:      stageRelation  $\leftarrow \text{StageOrder}[\text{current.stage}] - \text{StageOrder}[\text{prevAlert.stage}]$ 
13:      if stageRelation  $\geq 0$  then //阶段前

```

进或持续

```

14:         if current.timestamp >= prev-
Alert.timestamp then
15:             validPredecessors.append(pr-
ev-Alert)
16:         end if
17:     end if
18: end for
19: if validPredecessors  $\neq \emptyset$  then
20:     for each  $p$  in validPredecessors do
21:         if not HasEdge( $p$ , current) then
22:             AddEdge( $p$ , current)
23:         end if
24:     end for
25: end if
26: end for
27: origins  $\leftarrow$  GetZeroInDegreeNodes(graph)
28: for each origin in origins do
29:     AddEdge(source, origin)
30: end for
31: terminals  $\leftarrow$  GetZeroOutDegreeNodes
(graph)
32: for each terminal in terminals do
33:     AddEdge(terminal, target)
34: end for
35: graphs[ipPair]  $\leftarrow$  graph
36: end for
37: return graphs

```

(2) 频繁 1-项集挖掘

在为每个目标资产生成攻击图的同时，RASR 通过频繁 1-项集挖掘从告警数据中提取高频攻击类型。具体实现过程如算法 2 所示。算法的核心思想如下：首先，遍历所有告警数据，按照 (sIP, dIP) 进行分组，构建以 IP 对为单位的事务集合（行 1~5）；随后，统计每种攻击类型在所有事务中的出现次数，建立攻击类型到频率的映射关系（行 6~11）；最后，计算每种攻击类型的支持度，筛选出支持度大于等于最小支持度阈值的攻击类型作为频繁 1-项集（行 12~19）。

算法 2 频繁 1-项集挖掘算法

输入 告警集合 alerts，最小支持度阈值

minSupport

输出 频繁 1-项集 frequent 1-itemsets

```

1: transactions  $\leftarrow \emptyset$ 
2: for each alert  $a$  in alerts do
3:     pair  $\leftarrow (a.sIP, a.dIP)$ 
4:     transactions[pair]  $\leftarrow$  transactions[pair]
 $\cup \{a.alertType\}$ 
5: end for
6: attackCount  $\leftarrow \emptyset$ 
7: for each transaction  $T$  in transactions do
8:     for each alertType  $t$  in  $T$  do
9:         attackCounts[ $t$ ]  $\leftarrow$  attackCounts[ $t$ ] + 1
10:    end for
11: end for
12: frequent 1-itemsets  $\leftarrow \emptyset$ 
13: total  $\leftarrow |transactions|$ 
14: for each alertType  $t$  in attackCounts do
15:     if attackCounts[ $t$ ] / total >= minSupport
then
16:         frequent 1-itemsets=frequent 1-item-
sets  $\cup \{t\}$ 
17:     end if
18: end for
19: return frequent 1-itemsets

```

(3) 路径生成

基于上述两个步骤的输出，本文提出基于图遍历与双重筛选的路径生成方法。该方法通过深度优先搜索遍历攻击图获取所有候选路径，结合频繁 1-项集攻击类型匹配和高威胁告警检测两种筛选机制，确保提取的攻击路径既包含历史攻击模式中的常见告警类型，又能捕获关键的威胁事件。

具体步骤如算法 3 所示。首先，算法通过图遍历获取所有候选路径（行 1~3），检测其是否包含频繁告警类型（行 4~5），包含频繁告警类型的路径被判定为符合历史攻击模式。然后，算法识别路径中威胁等级超过预设阈值（本文设为 2 级）的关键告警（行 6）。需要说明的是，本文将攻击威胁程度按影响范围与破坏潜力划分为 1~5 级，具体见表 3。若阈值设为 1，则会纳入大量低风险误报，增加冗余；若阈值设为 3~5，则

会遗漏攻击链（攻击链强调攻击意图推进，攻击路径用于描述由告警推断的行为序列）早期的潜在风险路径；阈值设为2，可以精准覆盖攻击链萌芽阶段的关键节点，因此即使某条路径不包含频繁告警类型，只要包含攻击威胁程度 ≥ 2 的告警，仍将被纳入最终结果（行7~12）。该算法的时间复杂度为 $O(P \times L)$ ，其中 P 为路径数量、 L 为平均路径长度。

表3 攻击威胁程度级别划分

威胁程度	风险等级	关键特征
1	低风险	无实际攻击关联性
2	潜在风险	具备攻击链初始特征
3	中高风险	攻击链持续推进
4	高风险	攻击链成型并发起关键攻击
5	极高风险	攻击造成实质性损失

算法3 攻击路径生成算法

输入 攻击图集合 $graphs$ ，频繁 1-项集 frequent 1-itemsets

输出 攻击路径 $paths$

```

1:  $paths \leftarrow \emptyset$ 
2: for each graph  $g$  in  $graphs$  do
3:    $candidatePaths \leftarrow DFSAllPaths(g)$ 
4:   for each path  $p$  in  $candidatePaths$  do
5:      $hasFrequent \leftarrow any(alert.alertType \in frequent\ 1\text{-itemsets}\ for\ alert\ in\ p)$ 
6:      $hasHighThreat \leftarrow any(alert.degree > = threatThreshold\ for\ alert\ in\ p)$ 
7:     if  $hasFrequent$  or  $hasHighThreat$  then
8:        $paths \leftarrow paths \cup \{p\}$ 
9:     end if
10:  end for
11: end for
12: return  $paths$ 

```

3.3 场景增量重构

为实现攻击场景重构，需要对不同时间窗口生成的攻击路径进行组合分析。本文提出攻击路径增量融合算法，通过分析攻击阶段的时序关系和逻辑依赖关系，实现对动态网络环境中持续演进攻击的路径重构。

该算法的实现流程如算法4所示。对于每条增量路径，算法首先通过快速IP对匹配筛选出与已有路径具有相同“源-目的”IP对的候选路径（行1~4），以减少后续阶段关系分析的计算开销。针对IP匹配成功的路径对，算法将分析增量路径起点告警的攻击阶段与已有路径末端告警攻击阶段之间的时序关系（行10~12）。根据阶段分析结果，算法采用3种以下融合策略。

1) 直接拼接：若增量路径起始告警的攻击阶段不小于已有路径末尾告警的阶段，表明攻击处于正常推进或同阶段持续状态，算法将增量路径直接拼接到已有路径末端（行13~17）。

2) 路径插入：若增量路径起始告警的攻击阶段小于已有路径末尾告警的阶段，且已有路径存在与该告警类型相同的历史节点，则将增量路径插入该攻击路径（行8~22）。

3) 独立路径：若无法找到合适的融合位置，则将增量路径保留为独立的攻击路径（行5~9、行23~26）。

算法在所有融合过程中执行去重操作，既避免重复攻击类型的冗余存在，又可以保障攻击逻辑的完整性。最终，返回融合后的所有路径作为攻击场景重构的结果（行27）。该算法的时间复杂度为 $O(N \times B)$ ，其中 N 为新增路径数量、 B 为基线路径数量。在最坏的情况下，复杂度可能达到 $O(N \times B \times L^2)$ ，其中 L 为平均路径长度。然而在实际应用中，由于匹配路径数量有限且攻击路径长度较短，算法表现出良好的线性性能。

算法4 攻击路径增量融合算法

输入 已有路径集合 $basePaths$ ，增量路径集合 $newPaths$

输出 融合攻击路径 $fusedPaths$

```

1:  $fusedPaths \leftarrow basePaths$  //初始化融合结果
2: for each  $newPath$  in  $newPaths$  do
3:    $firstAlert \leftarrow newPath.start$ 
4:    $matchPaths \leftarrow FindIPMatchPaths(fusedPaths, firstAlert)$ 
5:   if  $matchPaths$  is empty then //独立路径
6:      $fusedPaths.add(newPath)$ 
7:      $fusionSuccess \leftarrow false$ 
8:   continue

```

```

9:  end if
10:  for each matchPath in matchPaths do
11:    lastAlert ← matchPath.end
12:    stageRelation ← StageOrder[firstAlert.
stage] - StageOrder[lastAlert.stage]
13:    if stageRelation >= 0 and firstAlert.
timestamp >= lastAlert.timestamp then //直接拼接
14:      AppendWithDeduplication(match-
Path, newPath) // 直接拼接
15:      fusionSuccess ← true
16:      break
17:    end if
18:    if FindHistoricalNode (matchPath,
firstAlert.alertType) then //路径插入
19:      BranchInsert(matchPath, newPath)
20:      break
21:    end if
22:  end for
23:  if not fusionSuccess then //独立路径
24:    fusedPaths.add(newPath)
25:  end if
26: end for
27: return fusedPaths

```

4 实验与结果分析

4.1 实验设置与数据集

本节对所提方法的攻击场景重构效果与性能进行评估。实验选取两个数据集进行验证，分别是 DARPA2000 LLDOS1.0 开源数据集^[22]和电力物联网真实攻击告警数据集，各数据集详情如下。

(1) DARPA2000 LLDOS1.0 数据集

该数据集由麻省理工学院林肯实验室采集，涵盖约 3 h 的网络攻击数据。该数据集具有公开可用性且包含完整的攻击标注，已被广泛用于攻击场景重构相关研究，因此本文选用该数据集评估本文方法的还原度。原始 DARPA2000 LLDOS1.0 数据集以网络流量（.dump 文件）形式提供，而 RASR 方法以入侵检测告警作为输入，因此本文首先将流量数据重放至部署了 Community Rules 和 Emerging Threats Rulesets 的 Snort 2.9 检测系统，生成对应的网络入侵告警数据。最终，共产生 217 条告警信息，对应一个完整的多阶段攻击

场景。该场景包含以下 4 个阶段。

1) 信息收集：攻击者探测运行了“sadmind”远程管理工具的主机。

2) 初始入侵：针对探测到的主机，多次利用 sadmind Remote-to-Root 漏洞发起攻击。

3) 据点建立：通过 Telnet 尝试登录目标主机，验证入侵是否成功。

4) 目标主机破坏：将可执行的二进制文件传输至受控主机，完成攻击落地。

(2) 电力物联网真实攻击告警数据集

电力物联网的网络安全设备持续产生大量的告警信息，对其进行实时分析有助于威胁事件的及时研判和处置。然而，由于真实环境中缺乏完整且清晰的数据标注，因此该数据集主要用于评估 RASR 方法的攻击场景重构效率，不适合用于评估 RASR 方法的还原度。

在数据预处理阶段，本文筛选出关键字段（包括 alertType、timeStamp、sIP、dIP、assetName、region、degree）均不为空的告警记录，最终用于实验的数据包含连续 9 个月内 17 个不同的检测系统上报的 55 646 条告警信息。这些告警源自 13 612 个攻击 IP 地址（其中有效的公网 IP 地址为 13 560 个）。

表 4 从攻击类型、攻击高峰时段、主要攻击源及所属互联网服务提供商（ISP）4 个维度对该数据集进行统计概括。各维度基于不同统计口径分别计算，彼此之间不存在一一对应关系。为突出数据的主要分布特征，表中每列数据均按本列统计量从高到低排序。

表 4 电力物联网告警数据集概要

攻击类型	攻击高峰时段	主要攻击源	所属 ISP
敏感文件探测 (30 046 次)	8:00 (12 348 次)	美国	美国国防部网络信息中心
注入攻击 (13 343 次)	9:00 (6 908 次)	英国	未知
密码猜解攻击 (3 454 次)	10:00 (4 580 次)	未知	美国有线电视公司
应用漏洞攻击 (2 972 次)	14:00 (4 056 次)	澳大利亚	微软公司
植入跨站脚本攻击 (XSS) 代码 (2 686 次)	15:00 (4 047 次)	日本	阿卡迈公司

4.2 还原度评估

为评估所提方法的有效性, 本节基于 DARPA 2000 LLDOS1.0 数据集开展攻击场景还原度实验, 选取完整性 (Completeness) 和健全性 (Soundness) 作为评估指标, 其计算公式分别如式(2)和式(3)所示。

$$\text{Completeness} = \frac{\#CCA}{\#RA} \quad (2)$$

$$\text{Soundness} = \frac{\#CCA}{\#CA} \quad (3)$$

其中, #CCA 表示正确关联的告警对数量, #RA 表示实际存在因果关系的告警对总数, #CA 表示算法输出的关联告警对数量。完整性反映了算法关联所有相关告警的能力, 健全性则衡量所关联告警对的准确程度。

应用本文方法对该数据集进行关联分析后, 共生成 18 条攻击路径, 路径详细信息如图 3 所示。图 3 中, 包含 3 个及以上节点的路径共 6 条, 可以完整描述整个攻击过程的路径共 3 条。图 3 和图 5 中告警信息包括: 蓝色表示的基于 UDP 协议的 RPC portmap sadmind 服务请求异常告警 (GPL RPC portmap sadmind request UDP)、黄色表示的深度使用 root 权限进行 sadmind 服务查询

的 RPC 攻击行为告警 (GPL RPC sadmind query with root credentials attempt UDP)、绿色表示的基于 rsh 服务的 root 权限远程登录攻击行为告警 (GPL MISC rsh root)、紫色表示的 TELNET 服务中检测到的非法登录尝试告警 (GPL TELNET Bad Login)、红色表示的检测到可执行文件文件格式 (ELF) 文件下载行为的策略类告警 (ET POLICY Executable and linking format (ELF) file download)。

为验证 RASR 的先进性, 本文选取 3 种经典且具有代表性的多阶段攻击场景重构方法与 RASR 进行对比。这些对比方法均以 IDS 警报为输入、输出事件级攻击场景。Poirot 系统^[9]采用因果分析方法挖掘告警之间的潜在关系, 并将其组织成告警链; 多步攻击告警关联 (MAAC)^[13]根据告警语义自动将簇分类并关联告警; 基于图的告警关联算法 (GAC)^[15]是一种基于图的告警关联算法, 通过对主机通信行为进行聚类并连接相应簇来实现攻击场景重构。表 5 中的还原度对比结果表明, RASR 在完整性和健全性方面均优于对比方法, 分别达到 71.43% 和 83.33%, 展示出其在发现告警关联关系方面的显著优势。

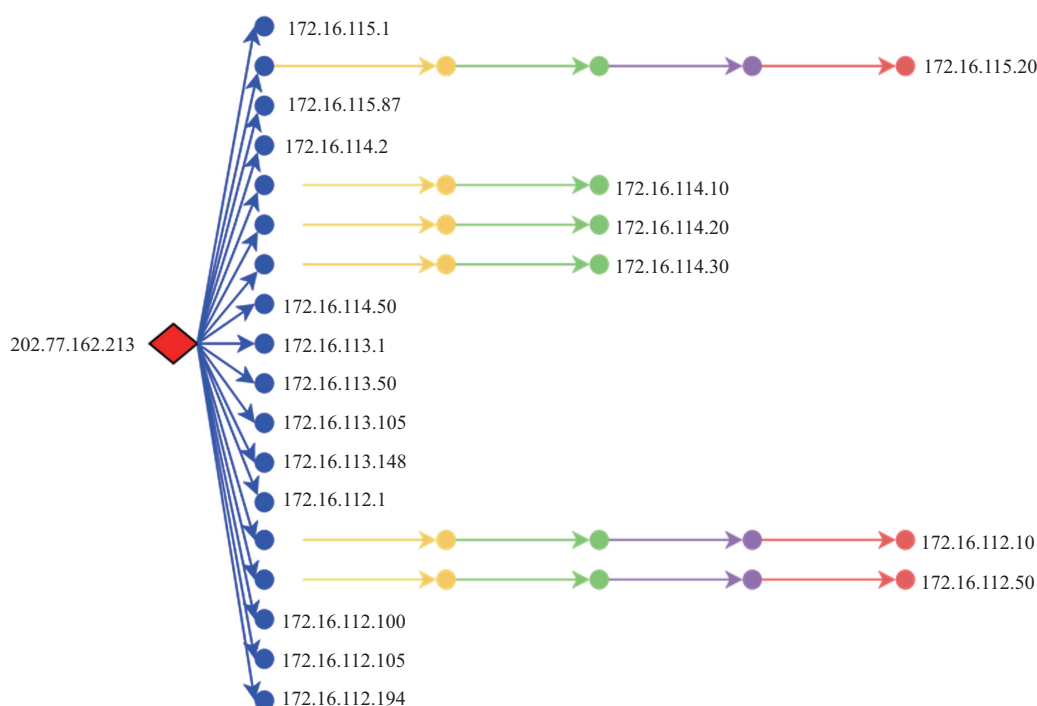


图3 RASR 基于 LLDOS1.0 数据集生成的攻击路径

表5 RASR与其他方法的还原度对比

方法	完整性	健全性
Poirot ^[9]	41.18%	4.46%
MAAC ^[13]	35.29%	40.00%
GAC ^[15]	64.71%	68.75%
RASR	71.43%	83.33%

4.3 效率评估

传统全量分析方法在每次新增告警到达时,需要重新分析所有告警数据。设已有告警数量为 M ,新增告警数量为 N ,生成的攻击路径数量为 P_{total} ,平均路径长度为 L ,则算法1的复杂度为 $O((M+N)^2)$,算法3的复杂度为 $O(P_{\text{total}} \times L)$,总体复杂度为 $O((M+N)^2)$ 。本文方法仅对新增告警进行增量处理,设基线路径数量为 B_{path} ,新增路径数量为 N_{path} ,新增告警关联分析(算法1)的复杂度为 $O(N^2)$,增量路径提取(算法3)的复杂度为 $O(N_{\text{path}} \times L)$,攻击路径增量融合算法(算法4)的复杂度为 $O(N_{\text{path}} \times B_{\text{path}} \times L^2)$ 。在实际应用中,由于 $N \ll M$ (新增告警远少于已有告警)、 $N_{\text{path}} \ll B_{\text{path}}$ (新增路径数量有限)及匹配路径数量有限且路径长度较短,因此总体复杂度近似为 $O(N \times M)$,相较于全量分析的 $O((M+N)^2)$ 具有显著优势。

本节基于电力物联网真实告警数据开展实验。根据第3.1节的分析,本文以24 h为时间窗口,选取某天作为第1天告警分析基线,对比了在告警数量逐日增加时,所提方法与全量分析方法的性能差异,运行时间对比如图4所示。

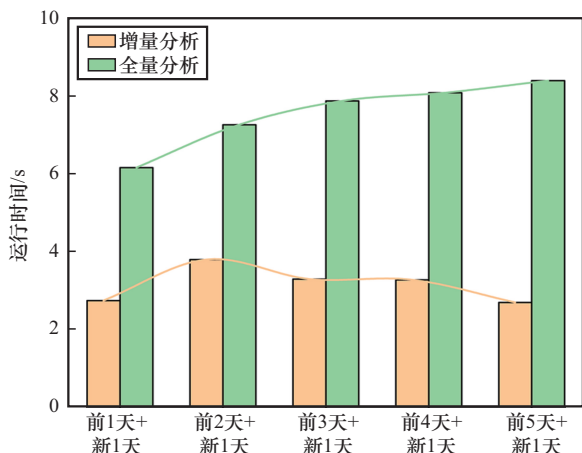


图4 逐日增量分析与全量分析的运行时间对比

实验以第1天的33 792条告警为基线,后续每天新增数百或数千条告警。实验结果表明,与传统的全量分析方法相比,本文方法在处理新增告警时运行时间显著缩短且保持稳定,而全量分析方法的运行时间随告警累积量线性增长。

4.4 案例分析

为更好地说明本文方法,本节结合DARPA2000 LLDOS1.0开源数据集和电力物联网真实数据集,给出两个告警攻击案例。

(1) DARPA2000 LLDOS1.0数据集案例

将DARPA2000 LLDOS1.0数据集的217条告警划分成前200条和后17条进行增量分析的实验。图5展示了详细的攻击路径生成过程和路径间增量融合后的结果。

在后17条告警中,首先出现了“GPL TELNET Bad Login”告警,说明202.77.162.213对172.16.112.50发起了Telnet尝试登录,但未成功;后续出现的sadmind告警,说明攻击者的漏洞扫描器仍在持续工作;最后出现“GPL MISC rsh root”和“ET POLICY Executable and linking format (ELF) file download”告警,说明攻击者已成功入侵并执行恶意文件传输。“GPL RPC portmap admind request UDP”对应的攻击阶段早于“GPL TELNET Bad Login”,RASR在攻击图中有效构建了攻击分支,并生成2条攻击路径。在路径融合阶段,RASR采用直接拼接等策略对分支路径进行融合,并对告警和路径进行去重操作,最终得到18条攻击路径。

(2) 电力物联网真实数据集案例

在电力物联网真实数据集上,基于历史告警数据已生成的攻击路径,本文对某天的新增告警进行了增量攻击场景重构。以其中一对“源-目的”IP的告警为例,该组告警包含敏感文件探测(信息收集阶段)、本地文件包含(初始入侵)和植入XSS跨站代码(初始入侵)3类安全事件。图6为在电力物联网真实数据集上的分析案例,展示了详细的路径生成与增量融合过程。具体而言,基于已有的告警序列,RASR生成了“本地文件包含”和“敏感文件探测→本地文件包含”

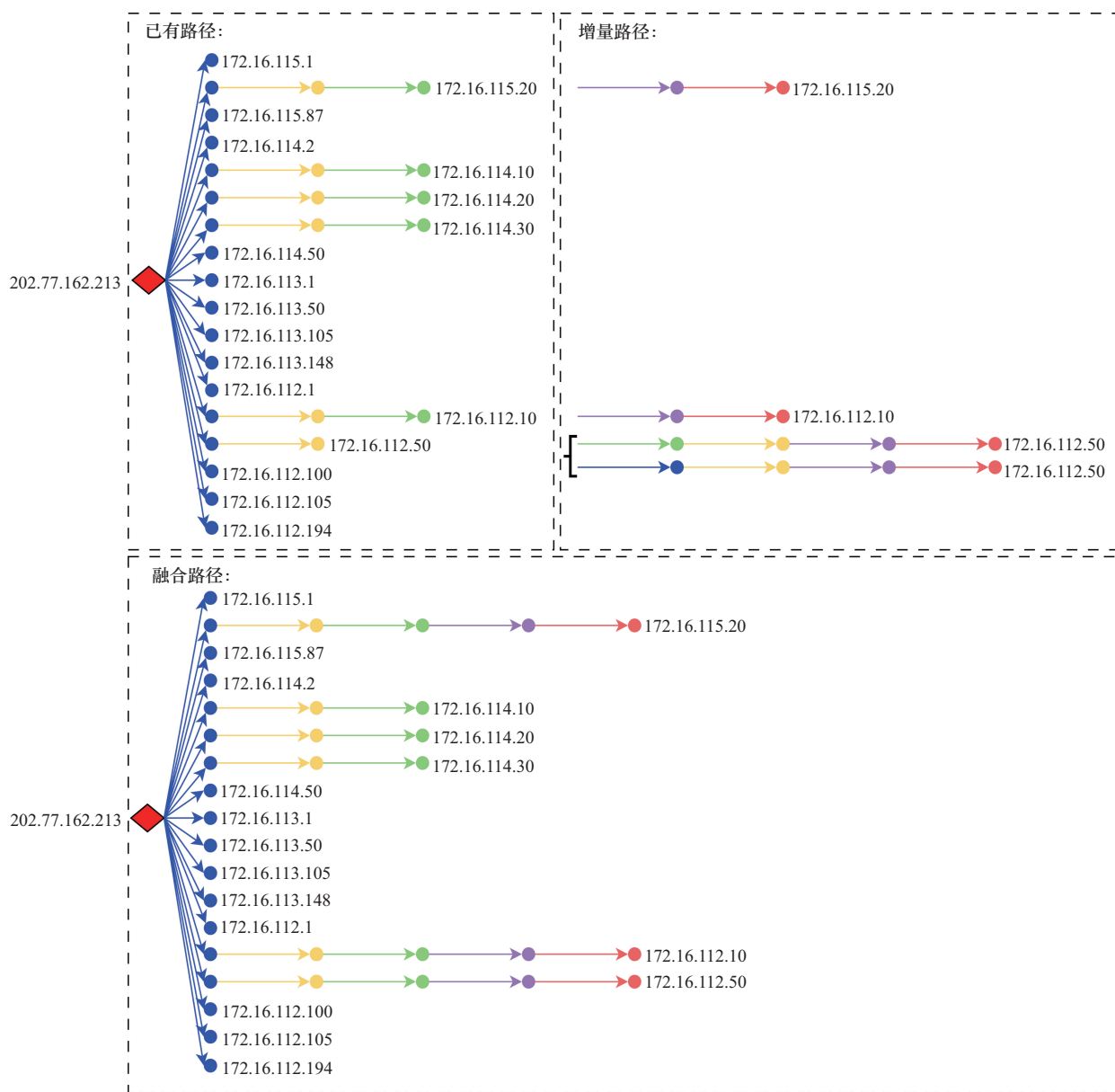


图5 在DARPA2000 LLDOS1.0数据集上的分析案例

2条攻击路径；基于对当天增量告警的分析，进一步生成“植入XSS跨站代码→本地文件包含”和“敏感文件探测→本地文件包含”2条新的攻击路径；在路径融合阶段，RASR通过直接拼接、告警和路径去重等操作，最终生成一条完整的攻击路径，实现了攻击场景的精准重构。从分析效率角度看，本案例中全量分析方法耗时13.33 s，而增量分析仅需1.04 s，效率提升约达12.82倍，显著体现了所提方法在实时性方面的优势。

颜色：

■ 敏感文件探测 ■ 本地文件包含 ■ 植入XSS跨站代码

已有路径：

26.47.X.X (Red Diamond) → 20.47.X.X (Yellow Circle) → 20.47.X.X (Yellow Circle)

新增路径：

26.47.X.X (Red Diamond) → 20.47.X.X (Yellow Circle) → 20.47.X.X (Yellow Circle)

融合路径：

26.47.X.X (Red Diamond) → 20.47.X.X (Yellow Circle) → 20.47.X.X (Yellow Circle)

图6 在电力物联网真实数据集上的分析案例

5 结束语

针对现有全量分析方法存在的结果滞后和效率瓶颈问题, 本文提出了一种基于告警增量分析的攻击场景实时重构方法。该方法通过预处理与窗口划分、攻击路径生成和场景增量重构等核心步骤, 实现了攻击场景的高效、实时重构, 为威胁事件的及时研判和应急处置提供了有效支撑。例如, 在企业网络安全环境中, 本文方法能够实现对海量告警数据的快速处理和攻击路径的动态追踪, 帮助安全运营中心及时识别复杂多步攻击路径, 有效缩短威胁响应时间。在未来的工作中, 可以探索引入大语言模型技术, 以更深入地识别和评估告警信息间的因果关联, 提升重构结果的可靠性。

参考文献:

- [1] 王郅伟, 何晞杰, 易鑫, 等. 基于 APT 活动全生命周期的攻击与检测综述[J]. 通信学报, 2024, 45(9): 206-228.
Wang Z W, He X J, Yi X, et al. Survey of attack and detection based on the full life cycle of APT[J]. Journal on Communications, 2024, 45(9): 206-228.
- [2] 胡源, 薛松, 张寒, 等. 近 30 年全球大停电事故发生的深层次原因分析及启示[J]. 中国电力, 2021, 54(10): 204-210.
Hu Y, Xue S, Zhang H, et al. Cause analysis and enlightenment of global blackouts in the past 30 years[J]. Electric Power, 2021, 54(10): 204-210.
- [3] 樊迪, 刘静, 庄俊玺, 等. 基于因果知识发现的攻击场景重构研究[J]. 网络与信息安全学报, 2017, 3(4): 58-68.
Fan D, Liu J, Zhuang J X, et al. Research on attack scenario reconstruction method based on causal knowledge discovery[J]. Chinese Journal of Network and Information Security, 2017, 3(4): 58-68.
- [4] 潘亚峰, 朱俊虎, 周天阳. APT 攻击场景重构方法综述[J]. 信息工程大学学报, 2021, 22(1): 55-60.
Pan Y F, Zhu J H, Zhou T Y. Survey on APT attack scenario reconstruction methods[J]. Journal of Information Engineering University, 2021, 22(1): 55-60.
- [5] Bryant B D, Saiedian H. Improving SIEM alert metadata aggregation with a novel kill-chain based classification model[J]. Computers and Security, 2020, 94(2): 101817.
- [6] Barzegar M, Shajari M. Attack scenario reconstruction using intrusion semantics[J]. Expert Systems with Applications, 2018, 108: 119-133.
- [7] Kenaza T, Aiash M. Toward an efficient ontology-based event correlation in SIEM[J]. Procedia Computer Science, 2016, 83: 139-146.
- [8] 王英州, 张耀方, 徐有方, 等. 工业控制系统物理层报警量关联算法[J]. 网络与信息安全学报, 2025, 11(1): 178-188.
Wang Y Z, Zhang Y F, Xu Y F, et al. Physical layer alarm correlation algorithm for industrial control system[J]. Chinese Journal of Network and Information Security, 2025, 11(1): 178-188.
- [9] Yang J, Zhang Q, Jiang X F, et al. Poirot: causal correlation aided semantic analysis for advanced persistent threat detection[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(5): 3546-3563.
- [10] Zang X D, Gong J, Zhang X C, et al. Attack scenario reconstruction via fusing heterogeneous threat intelligence[J]. Computers and Security, 2023, 133: 103420.
- [11] 张玉臣, 孙澄, 姜迎畅, 等. 融合威胁情报与知识图谱的网络攻击溯源方法[J]. 情报杂志, 2024, 43(8): 72-83, 91.
Zhang Y C, Sun C, Jiang Y C, et al. A traceability method of network attacks combining threat intelligence and knowledge graph[J]. Journal of Intelligence, 2024, 43(8): 72-83, 91.
- [12] Zhao J, Tang T, Bu B, et al. Graph neural network-based attack prediction for communication-based train control systems[J]. CAAI Transactions on Intelligence Technology, 2024: cit2. 12288.
- [13] Wang X Y, Gong X R, Yu L, et al. MAAC: novel alert correlation method to detect multi-step attack[C]//Proceedings of the 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Piscataway: IEEE Press, 2021: 726-733.
- [14] Wang W B, Yi P, Jiang J F, et al. Transformer-based framework for alert aggregation and attack prediction in a multi-stage attack[J]. Computers and Security, 2024, 136: 103533.
- [15] Haas S, Fischer M. GAC: graph-based alert correlation for the detection of distributed multi-step attacks[C]//Proceedings of the 2018 ACM Symposium on Applied Computing. New York: ACM, 2018: 979-988.
- [16] Hu H, Liu J, Zhang Y C, et al. Attack scenario reconstruction approach using attack graph and alert data mining[J]. Journal of Information Security and Applications, 2020, 54: 102522.
- [17] Jia B W, Tian Y Z, Zhao D, et al. Bidirectional RNN-based few-shot training for detecting multi-stage attack[C]//Proceedings of the 16th International Conference on Information Security and Cryptology. Berlin: Springer, 2021: 37-52.
- [18] Farhadi H, Amirhaeri M, Khansari M. Alert correlation and prediction using data mining and HMM[J]. The ISC International Journal of Information Security, 2011, 3(2): 77-101.
- [19] 吕磊, 刘家宇, 李琦, 等. 基于改进序列模式挖掘算法的告警关联模型[J]. 电讯技术, 2023, 63(6): 911-917.
Lyu L, Liu J Y, Li Q, et al. Alert association model based on improved sequence pattern mining algorithm[J]. Telecommunication Engineering, 2023, 63(6): 911-917.
- [20] Mao B, Liu J, Lai Y, et al. MIF: a multi-step attack scenario reconstruction and attack chains extraction method based on multi-information fusion[J]. Computer Networks, 2021, 198: 108340.
- [21] ATT&CK v19. [EB]. 2026.
- [22] Defense Advanced Research Projects Agency (DARPA). 2000 DARPA intrusion detection evaluation data set[EB]. 2025.

[作者简介]



赵新建（1987-），男，国网江苏省电力有限公司信息通信分公司高级工程师，主要研究方向为电力数字化和网络安全。



张玉健（1984-），男，东南大学副教授，主要研究方向为分布式系统和系统安全。



汤慧敏（2000-），女，东南大学硕士生，主要研究方向为网络安全和攻击溯源。



张颂（1996-），男，国网江苏省电力有限公司信息通信分公司工程师，主要研究方向为电力数字化和网络安全。



陈石（1995-），男，国网江苏省电力有限公司信息通信分公司工程师，主要研究方向为电力数字化和网络安全。



程光（1973-），男，东南大学教授，主要研究方向为流量分析和网络架构。