南京郵電大學
Nanjing University of Posts and Telecommunications

基于威胁情报知识图的攻击假设生成器
Attack Hypotheses Generation Based on Threat Intelligence Knowledge Graph

汇报人：孙艺博

汇报时间：2024.6.14

IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 20, NO. 6, NOVEMBER/DECEMBER 2023

CONTENT

01

研究背景

研究背景

现在的网络攻击行为已经由传统单一简单的攻击机制转变为多阶段多步骤复杂的攻击机制。

例如：Zeus僵尸网络通过扫描检测、溢出攻击、感染目标、传播病毒及窃取数据五个攻击步骤发动攻击，这会造成严重且恶劣的影响。

通过对攻击者使用的工具和攻击模式提供更深入的了解，过去攻击的网络威胁情报可能有助于攻击重建和正在进行的攻击过程的预测。因此，网络安全分析师使用威胁情报，告警关联，机器学习和高级可视化来产生合理的攻击假设，进而推断出攻击者可能的后续攻击步骤，实现及时发现主动防御。这种方法将成为未来的趋势。

02

相关知识与动机

相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

CTI是结构化的、可操作的信息，用于识别对手及其动机、目标、能力、资源和战术。它包括以可测量事件和事件解释背景为形式的循证知识。这些信息可以归为四类:(i)技术、(ii)战术、(iii)操作和(iv)战略。

从CTI中提取的信息提高了分析人员识别相关威胁并及时做出反应的能力。因此CTI是提高入侵检测、事件响应、实时分析、取证调查和威胁搜索等各种安全解决方案效率的有力手段。

此外，这些低级攻击工件(技术CTI)可以快速操作; 然而，很可能在短时间内就过时了。因此，低级CTI(例如ioc)对于制定攻击假设的价值是有争议的。

Ioc（ Indicator of Compromise ）： 威胁迹象。这些指标通常是指在网络安全中用于检测或确认是否发生了安全事件或受到了攻击的特定指标，如恶意文件的哈希值、恶意 IP 地址、异常网络流量模式等。

故有必要将更抽象、更健壮(就过时而言)的战术和技术(高级CTI)与可操作的ioc(低级CTI)的好处结合起来。

相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

IOA （Indicator of Attack）：攻击迹象，与 IOC 类似，IOA 用于指示系统可能受到攻击的特定行为或模式。与 IOC 不同，IOA 更专注于攻击的行为特征，而不仅仅是已知的标识符。

在文献中，IoA被定义为攻击中可用的CTI的整体，包括战术、技术和程序(TTP)的高级描述。

一些常见的 IOA 包括：

异常的进程执行： 恶意软件可能启动的进程，与正常操作不符。

不寻常的文件访问行为： 对系统关键文件的异常读取或写入。

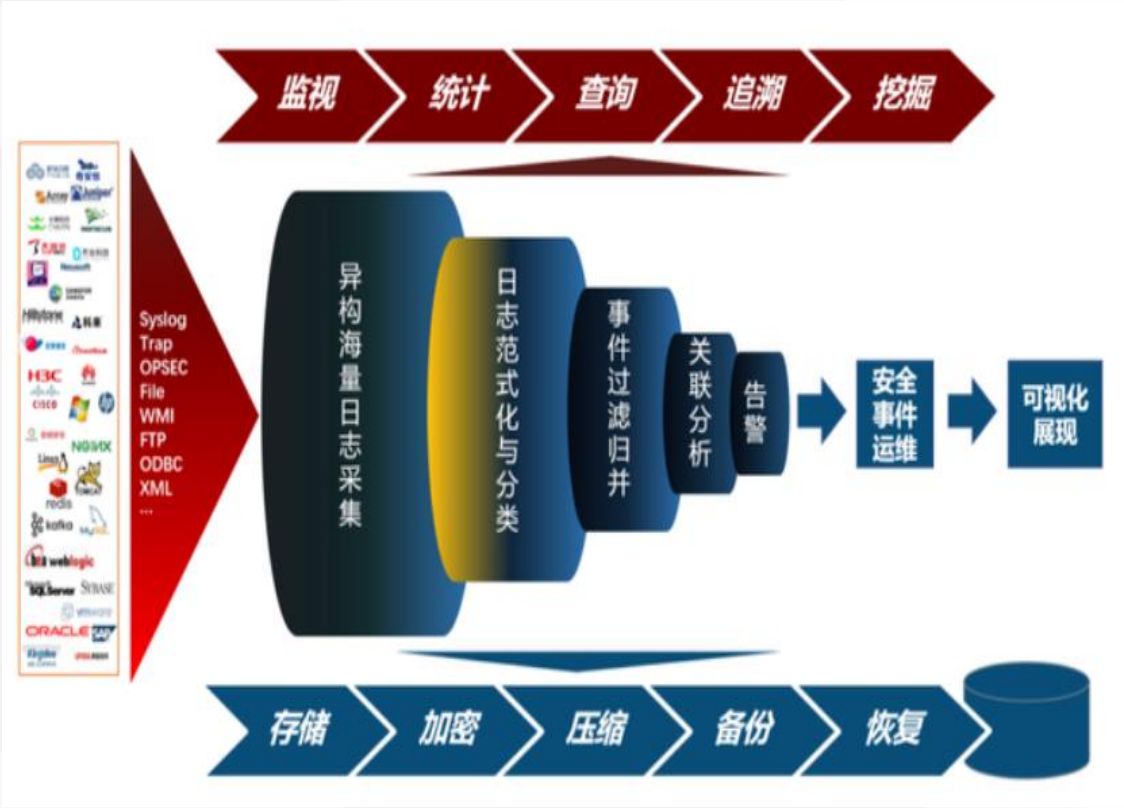
由于没有组织能够通过记录攻击工件来完全了解威胁情况，因此CTI的重要性在于它能够以机器对机器的方式在合作伙伴之间共享。通过共享谁，什么，在哪里，如何，以及何时恶意活动，组织。获得威胁形势的整体视图，从而提高其网络安全准备。

然而，CTI的共享给CTI的可信度带来了新的风险。因此，评估CTI的正确性和可靠性至关重要。

相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

SIEM代表安全信息与事件管理（Security Information and Event Management）。SIEM系统是一种综合性的安全解决方案，用于实时监控和分析组织内的安全事件和活动。这类系统通过收集、聚合、分析和报告来自各种网络设备、操作系统和应用程序的日志和事件数据，帮助组织识别和响应潜在的安全威胁。



相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

此外，共享CTI的一个主要挑战是，它以各种不同的格式从各种来源共享。

因此，信息共享需要流程化和结构化为了形成一种共享CTI的标准语言，美国国土安全部网络安全和通信办公室向MITRE提供了资金，以开发结构化威胁信息表达(STIX)语言。

STIX涵盖了网络安全概念的整个范围，包括observables, IoCs, attack patterns, tools, malware, threat actors, courses of action 等等。

STIX元素被称为STIX域对象（STIX Domain Object），SDO：用于表示威胁信息的各个方面，如攻击模式、威胁行为者、恶意软件等。

将SDO中的元素分为两种：

低级CTI：如可观察对象和ioc。

高级CTI：如攻击模式、工具和威胁参与者。

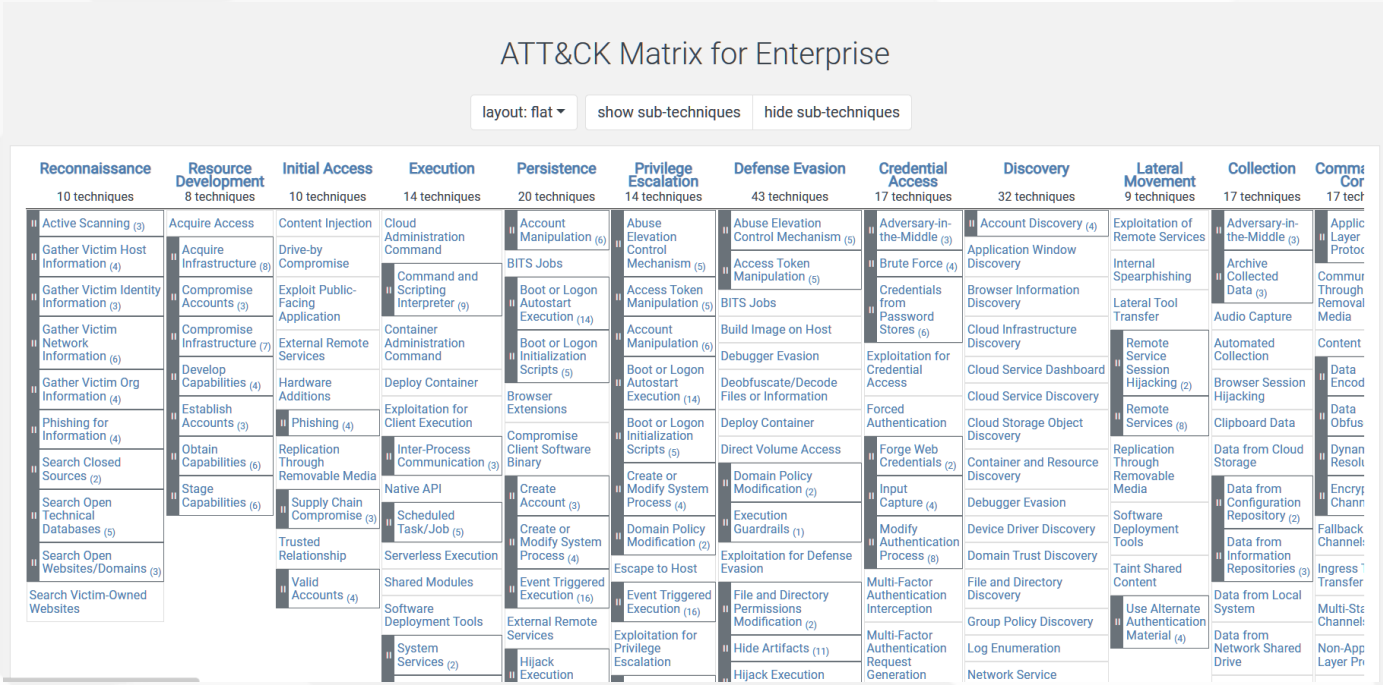
相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

在网络安全领域，MITRE 还提供了许多重要的贡献，其中包括：

ATT&CK （Adversarial Tactics，Techniques，and Common Knowledge）框架：MITRE ATT&CK 框架是一个广泛使用的网络攻击知识库，用于描述攻击者的战术、技术和程序（TTPs），并提供了一种通用的语言来描述和分析网络攻击行为。该框架包含了18种战术。

该框架分为三个技术域：Enterprise，Mobile，ICS。在这里我们主要关注Enterprise。



相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

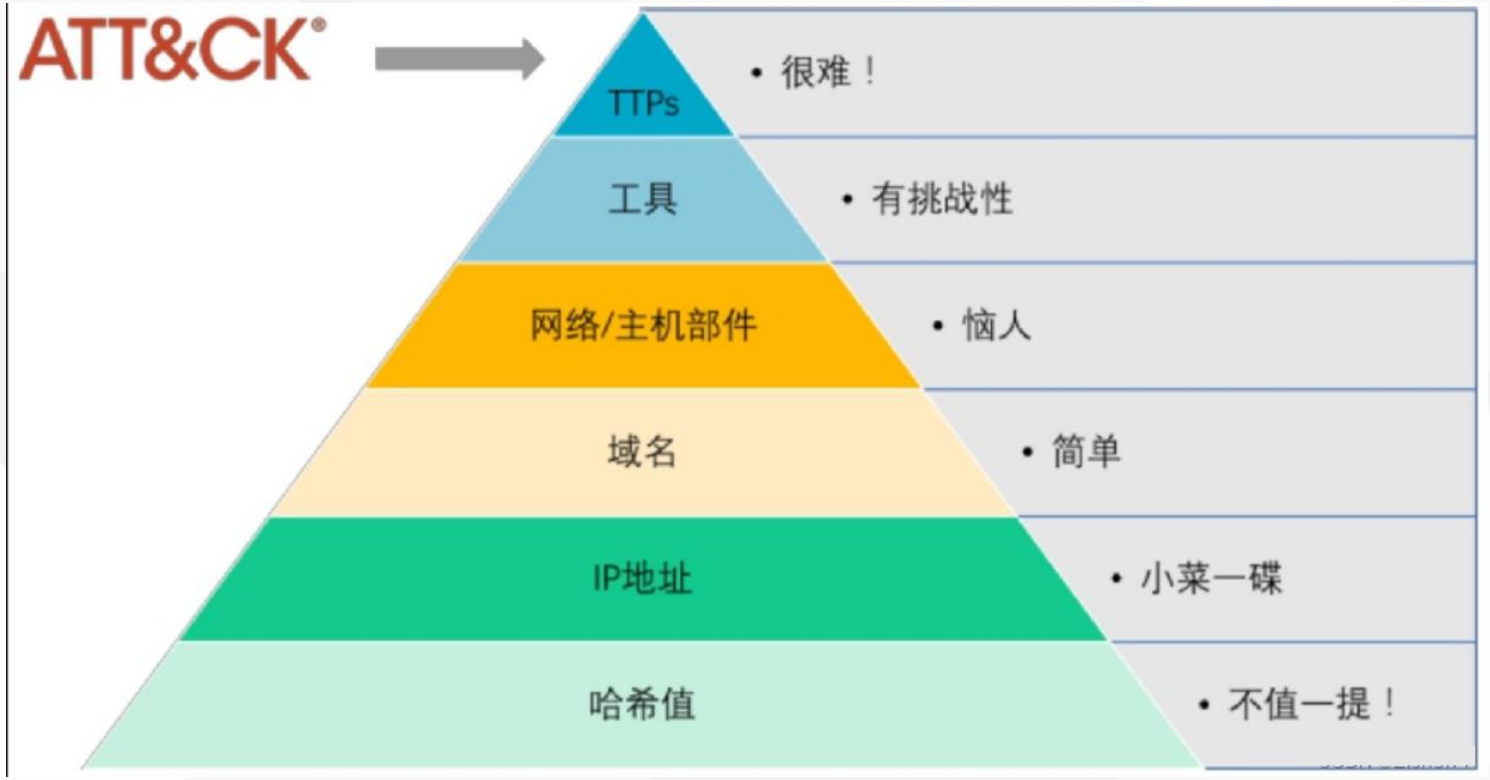
ATT&CK观察的主要目标是攻击者的TTPs，并将观察结果记录进行总结分类，从而形成知识库。

相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

ATT&CK中痛苦金字塔概念。

痛苦金字塔模型由IoC组成，通过IoC进行组织分类并描述各类IoC在攻防对抗中的价值。



相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

还有一些其他公司开发的攻击框架如：

Cyber Kill Chain：Cyber Kill Chain是由Lockheed Martin开发的攻击链框架，确定了攻击者必须完成什么才能实现他们的目标，其展示的七个步骤增强了对于攻击的可见性，并丰富了分析师对攻击者TTPs（Tactics, Techniques and Procedures,战术、技术和程序）的了解。

相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

二者区别：

网络杀伤链可以说是一个描述性框架。它的详细程度远不如 MITRE ATT&CK。与 MITRE ATT&CK 的 18 种策略（包括仅限移动的策略和仅限 ICS 的策略）相比，它仅涵盖七种策略。它没有为针对移动或 ICS 平台的攻击提供离散模型。它编入目录的内容也没有 MITRE ATT&CK 中的策略、技术和程序信息那么详尽。

另一个重要区别是，网络杀伤链基于以下假设：任何网络攻击必须依次完成对抗策略才能获得成功，阻止任一策略都会“中断杀伤链”并阻止对手实现其最终目标。MITRE ATT&CK 并不采用这种方法；它专注于帮助安全专业人员在遇到任何情况时识别并阻止或缓解各种对抗策略和技术。

相关知识及动机

网络威胁情报（Cyber Threat Intelligence）：

之前的研究，虽然现有的CTI本体、语言和存储库，但它们不足以有效地生成攻击假设。本文延续之前的研究并进行深化。

在本文中作者提出一个综合多层次威胁知识库AttackDB，它结合多个开源威胁情报来源，并糅合了各种恶意软件使用的不同技术的整体视图，有效地将它们与低级CTI连接起来。

相关知识及动机

威胁搜寻（Threat Hunting）：

威胁搜寻分为两类：

一方面，一些专家将威胁搜寻定义为主动寻找可能正在进行的攻击的早期迹象，而不是等待警报表明可疑活动。

另一方面，威胁搜寻可能指的是响应警报而启动的调查过程。此过程可能包括高级分析、取证调查、目标数据收集或策略更新。

主动和被动的威胁搜寻的主要区别在于调查的触发。主动威胁搜索依靠CTI主动搜索潜在的恶意行为。被动威胁搜寻包括法医调查和攻击假设测试，以响应指示此类行为的警报。

虽然大多数先前的威胁搜索研究都集中在自动检测和响应上，但目前的工作主要集中在威胁搜索过程的假设生成阶段。

相关知识及动机

假设生成（Hypothesis Generation）：

1. 运用运营和战略情报进行推理：

攻击重建通常是成功的威胁搜索过程的输出，它指的是通过呈现攻击者成功执行的不同步骤来描述威胁。安全分析师应该能够通过指出基于收集的证据及其分析的相关事件来解释每个步骤是如何实现的。

前人的工作如下：

（1）因果攻击图：

Milajerdi等人[32]采用因果来源图与攻击者活动一起对组织结构和流程进行建模。他们的目标是从系统日志中推断高级ttps。从日志中推断出用于构造起源图的因果关系需要在目标环境中进行广泛的威胁模拟。这些信息通常不是主要CTI来源提供的。相比之下，本文中提出的TTP推断依赖于各种来源发布的现成的通用CTI。

（2）攻击检测：

Bhatt等人[35]也提出了一种威胁检测模型。所提出的模型用于改进对给定相关事件的持续攻击的假设，以及对网络杀伤链(Cyber Kill Chain，CKC)的了解。与此类别中的工作相反，AHG不是用于攻击检测的。相反，我们的目标是在假设组织已经受到攻击的情况下，根据一组可疑的工件推断出最可能的攻击技术。

相关知识及动机

假设生成（Hypothesis Generation）：

1. 运用运营和战略情报进行推理：

（3）从工件推断TTPs/CKC阶段：

工件：通常指的是在进行调查、研究或分析过程中收集到的各种证据、信息或数据。在网络安全领域中，工件可以是指网络流量数据、日志记录、恶意软件样本、攻击者使用的工具、系统配置信息等。

Wang[37]提出了一种威胁搜寻假设开发方法，用于识别威胁参与者、目标资产、相关漏洞和工件。他们建议使用CTI的探索性数据分析来生成初始假设和验证假设。

Giura等人[38]提出了一种攻击金字塔，旨在通过CKC阶段(以金字塔的层次表示)和组织的环境(以金字塔的平面表示)，如物理、网络、用户和应用程序来捕获攻击者的运动。攻击者的目标位于金字塔的顶端，可以通过从一个事件跳到另一个事件来实现。金字塔可以检测攻击路径和攻击者的目标。

相关知识及动机

假设生成（Hypothesis Generation）：

1. 运用运营和战略情报进行推理：

Iqbal等[20]通过从文本报告中提取实体，为CKC和Pyramid of Pain模型构建了统一的图表示。他们比较了同一攻击的两种变体，并表明他们可以使用图来映射攻击的每个阶段的TTP，这将允许分析师依靠作者提供的方法从图中推导和预测缺失的TTP。然而，所提出的方法不提供自动化，并且仅展示了两种恶意软件的预测能力。

总的来说，这些开创性的工作构成了作者建立假设生成方法的基础。大多数先前的最先进的方法使用基于专家的规则来演示对TTPs的推断，这些规则仅针对几种表现出类似攻击流的攻击。作者构建了AttackDB——一个自上而下构建的综合知识图谱，依靠主要的CTI资源(MITRE ATT&CK、AlienVault、X-Force Exchange和VirusTotal)。AttackDB能通过依赖基于网络的推理而不是手动定义规则集来简化TTP推理过程。

相关知识及动机

假设生成（Hypothesis Generation）：

2. 运用威胁情报知识图谱进行推理：

在网络安全领域，知识图的使用尚处于早期阶段，主要用于可视化，较少用于预测。

最早使用知识图谱进行分析的研究之一是由Lee等人[40]完成的，他们从开源智能中构建了一个知识图谱。基于已建立的图算法，改进了恶意节点和攻击基础设施的识别，改进了攻击组之间的关系和相似度。特别是，他们使用页面排名和间隔度量来检测图中的相关信息。

受这些工作的启发，作者采用监督机器学习和链接预测方法，通过知识图分析从工件中推断TTP。

作者旨在为自动推断IoA，以及人工分析员推断的IoA的细化做出贡献。与上述提到的图形对齐和逻辑规则相比，本文提出的方法可以完全自动推断TTP。作者依赖包含许多攻击变体并且易于安全团队获取的CTI来源。作者在本文中提出的威胁情报知识库AttackDB具有独特性，包含了来自MITRE ATT&CK的所有恶意软件家族，并将高级TTP与低级可观察工件进行了链接。

03

方案设计

方案设计

1.多层次威胁知识库：

（1）结构

AttackDB包含痛苦金字塔的所有层次的SDO，从抽象概念(如战术和顶级技术)到IOC和特定的可观察对象(如哈希、互联网协议(IP)地址和域名)。

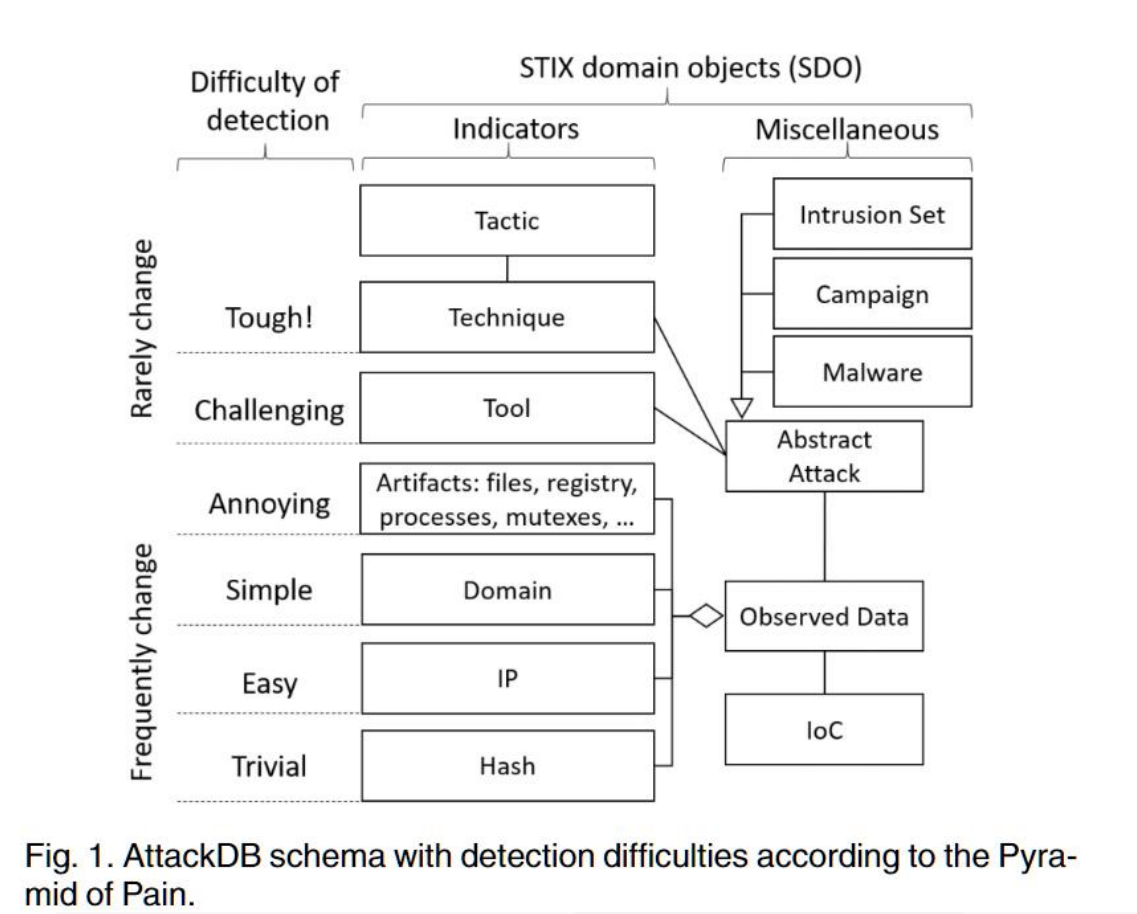


Fig. 1. AttackDB schema with detection difficulties according to the Pyramid of Pain.

方案设计

1.多层次威胁知识库：

（2）数据融合

在当前的实现中，AttackDB知识图存储在Neo4j2数据库中。通过AttackDB构建了一个丰富的知识库，该知识库由MITRE ATT&CK企业知识库、OTX、X-Force和VirusTotal组成。不同对象之间的关系包括在内，以构建知识图，因为它们在从不同CTI源提取的恶意软件分析报告中进行了描述。关于数据提取过程和数据融合的详细信息如下。如图2所示。

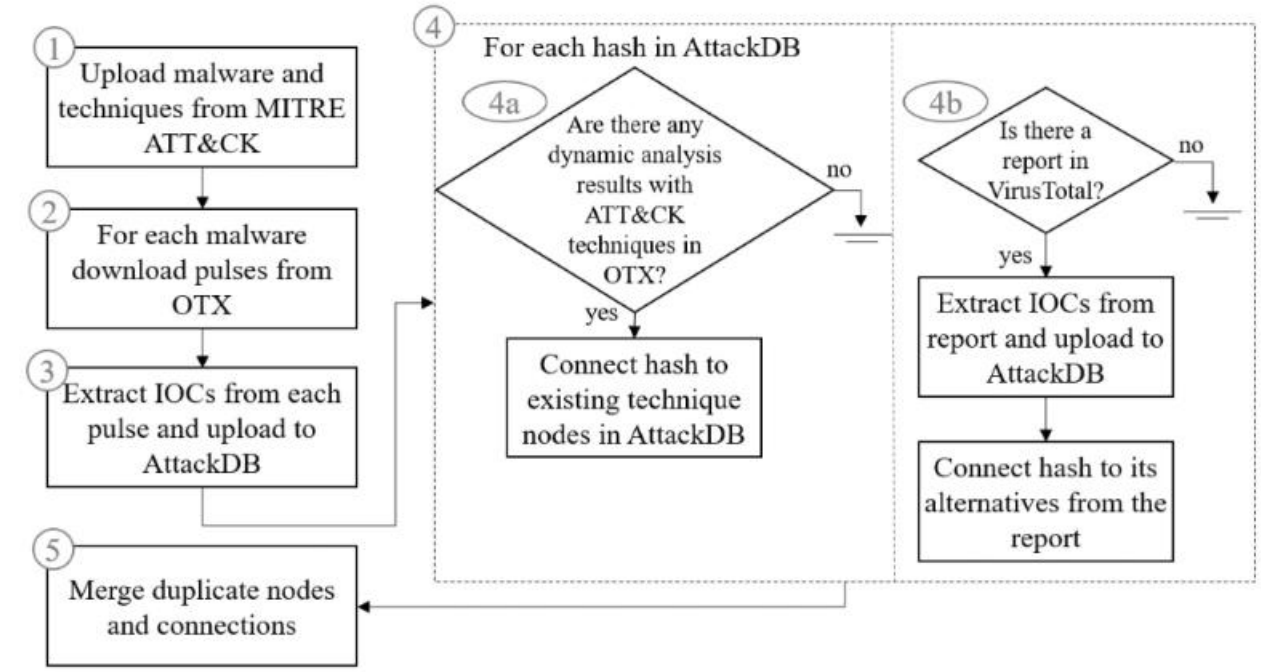


Fig. 2. Flow chart of AttackDB's construction.

方案设计

1.多层次威胁知识库：

AttackDB最终结构

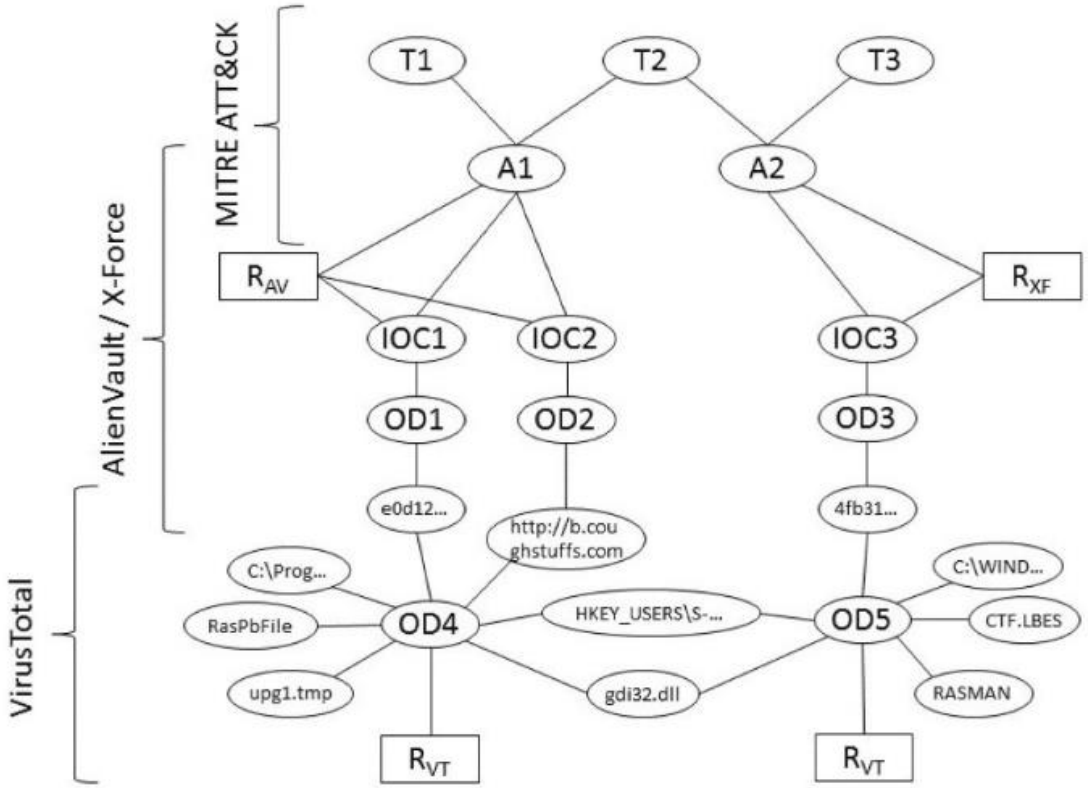


Fig. 3. An illustration of AttackDB's structure.

方案设计

1.多层次威胁知识库：

（3）重复数据和恶意软件别名

行为工件可能通过多个路径连接到恶意软件。当VirusTotal分析了同一恶意软件的多个实例时，就会发生这种情况。

在生成的 AttackDB 中，有连接到两个或多个恶意软件节点的散列节点。由于哈希节点表示特定的恶意软件实例，与恶意软件节点的模糊连接需要进行额外的处理。

这种现象的可能原因如下所述：

(1)报告描述了一些因素，如参与者、活动、同一恶意软件家族、感染机制或ioc重叠，这些因素在多个恶意软件实例中很常见;

(2)描述单一特定恶意软件的报告。第一种类型的报告通常包含对多个恶意软件二进制文件的引用。在这种情况下，攻击节点不仅连接到其代表性哈希，而且连接到恶意软件分析报告中提供的所有哈希。第二种类型的报告描述了包含单个代表性散列的单个恶意软件，但是可能有不同的报告以不同的名称描述相同的恶意软件。在这种情况下，一个散列连接到报告中出现的所有恶意软件别名。

方案设计

1.多层次威胁知识库：

（4）知识库概述

由此产生的攻击数据库融合了来自1,675个AlienVault脉冲、281个IBM X-Force报告和53,005个VirusTotal报告的数据。包含253个恶意节点，144216个iocs。大约有6万个iocs是文件哈希，其余的是域名、ip等。AttackDB中总共有超过50万个可观察对象。图中的所有技术都与某些恶意软件有关。每个恶意软件的平均技术数量是10.3。在本研究中，我们只包括与MITRE恶意软件相关的技术。

方案设计

2.攻击假设生成器：

假设有一个前面所描述的知识图。用于表示攻击的恶意软件SDO被分组在一个表示为A的超集中。攻击描述性SDO包含有关攻击所使用或攻击目标的元素的信息。可用于描述攻击的SDO，特别是技术(又称攻击模式)、IOC、观察到的数据和特定的可观察对象，被分组在一个超集中，表示为D。

网络安全知识图谱：网络安全知识图谱KG=<A,D,R>是一个图，其中A包含表示过去攻击的节点，D包含描述节点(具体来说，技术、IOC、观察数据和可观察对象)。R是所示模式连接相关SDO的有向链路的集合。

攻击描述：根据攻击表示 “a” 到 “A”，作者将距离 “a” 最多五个跳跃的集合 “v”，表示为攻击描述 AD_a= {v | dist(a, v) ≤ 5} ⊆ D。并将攻击技术表示为AT_a = {v | v ∈ AD_a ∧ type(v) = Technique}。

方案设计

2.攻击假设生成器：

假设组织中正在发生可疑事件。AHG的目标是对可能的攻击过程提出假设。图4提供了假设生成过程的概述。 生成的假设包括一组与观察数据和彼此密切相关的MITRE ATT&CK技术。

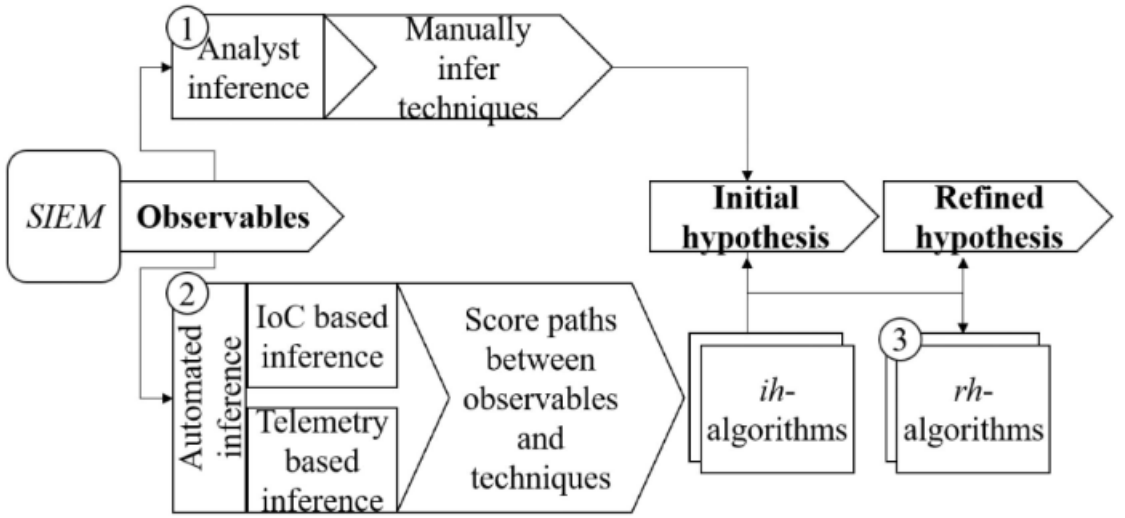


Fig. 4. Overview of the hypothesis generation process.

方案设计

2.攻击假设生成器：

（1）初始攻击假设生成

给定一个知识图谱 KG 和 SIEM 中的 COD（当前观测数据），生成一个初始的攻击假设，表示为 $AT_{\mathrm{a\_}new}^{\mathrm{init}}$，该假设由与 COD 密切相关的技术组成。

作者采用几种方法来生成将 COD 映射到技术的关系:

a：TF-IDF

b：朴素贝叶斯推理

c：多层朴素贝叶斯推理

初始假设生成可以基于模拟人类分析员（策略标记为 H）或自动化进行。COD 可能是 IoC 或 telemetries，这取决于所采用的推断技术方法。

方案设计

2.攻击假设生成器：

（1）初始攻击假设生成

a. TF-IDF: （词频-逆文档频率）

基于词频-逆文档频率（TFIDF）的技术评分机制，并将其用于评估技术与观察数据之间的相关性。

1. TFIDF基本概念：TFIDF是信息检索领域常用的一种技术评分机制。在这种情况下，技术类似于文档，而可观察的数据类似于搜索项，用于TFIDF计算。

2. 观察数据与技术的关联：对于给定的观察数据obs和技术t，如果obs出现在使用t的攻击报告中，则认为t与obs相关。一个观察数据可能出现在多个报告中，并且通过多个攻击与一个技术相关联。

3. 连接观察数据到技术的路径：定义了连接观察数据到技术的路径的数量，记为TF(obs, t)。

具体而言：

- 当技术的评分基于IoC时，路径的集合包括从obs到t的路径。
- 当技术的评分基于telemetries时，路径的集合包括从obs到t的路径。

方案设计

2.攻击假设生成器：

（1）初始攻击假设生成

a. TF-IDF: （词频-逆文档频率）

4. 逆文档频率（IDF）：IDF衡量了技术的普遍性，即在AttackDB中有多少技术与给定的观察数据相关。这里采用了最简单的IDF评估方法，即总技术数量的对数除以AttackDB中与给定观察数据相关的技术数量。

5. TFIDF评分：技术t的TFIDF评分是其TF值与IDF值的乘积。t的总评分是所有COD的TFIDF值的总和。通过这个TFIDF评分机制，可以确定技术与观察数据之间的相关性，以及技术的普遍性，从而帮助生成与潜在攻击行为相关的假设。

通过这个TFIDF评分机制，可以确定技术与观察数据之间的相关性，以及技术的普遍性，从而帮助生成与潜在攻击行为相关的假设。

方案设计

2.攻击假设生成器：

（1）初始攻击假设生成

b. 朴素贝叶斯推理

- 多项式朴素贝叶斯分类器（NB-C）：采用基于向量计数的多项式NB-C。多项式NB-C可以基于词向量计数或TFIDF进行建模。这里介绍了一种基于向量计数的NB-C。
- 先验概率：先验概率是从知识图谱（KG）中提取的，反映了与观察数据相关的先验知识。在这里，先验概率是根据相关可观察数据的数量进行提取的。
- 计算后验概率
- 计算p（obs|t）
- 计算p（COD）

$$P(t|COD) = \frac{P(t) \cdot \prod_{obs \in COD} P(obs|t)}{P(COD)}$$

$$P(obs|t) = \frac{TF(obs,t) + \alpha}{\sum_{t \in T} TF(obs,t) + \alpha}$$

$$P(COD) = \prod_{obs \in COD} P(obs),$$

方案设计

2.攻击假设生成器：

（1）初始攻击假设生成

b. 朴素贝叶斯推理

平滑先验：使用平滑先验参数a，可以进行拉普拉斯平滑（a = 1）或Lidstone平滑（a < 1）。平滑先验可用于处理不在AttackDB中的obs和连接，同时防止概率为零。使用这个平滑参数可能会提高分类的准确性。

通过这种朴素贝叶斯推断的方法，可以根据观察数据生成与潜在攻击相关的初始假设，并考虑到技术和观察数据之间的相关性。

方案设计

2.攻击假设生成器：

（1）初始攻击假设生成

c. 多层朴素贝叶斯推理

多层朴素贝叶斯推断（Multi-Layer Naïve Bayesian Inference）的方法，利用了知识图谱的本体。该推断算法允许利用AttackDB中的因果关系，因此逐步降低了NB-C的条件独立性假设。该方法包括两个层次（见图5）：

(I) 第一层：从恶意软件推断技术的NB-C。

(II) 第二层：从COD推断恶意软件的NB-C。

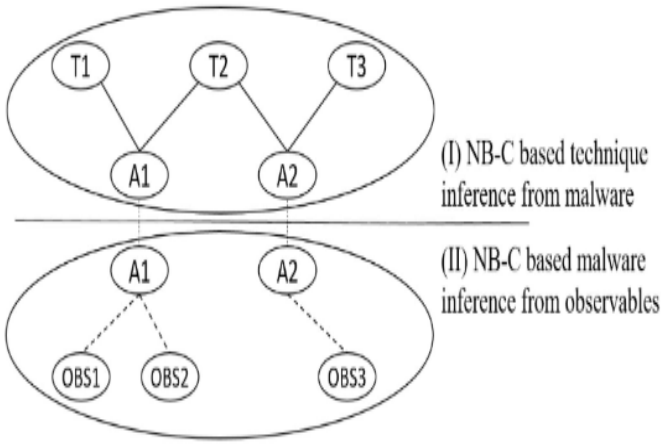


Fig. 5. Framework of multi-layer naïve Bayesian inference of initial hypotheses within AttackDB.

方案设计

2.攻击假设生成器：

（1）初始攻击假设生成

c. 多层朴素贝叶斯推理

根据总概率法则，给定COD时，技术t的概率可以定义如下：

$$P(t|COD) = \sum_{a \in \mathbb{A}} \frac{P(t) \cdot P(a|t) \cdot P(a|COD)}{P(a)}$$

具体而言，计算过程如下：

- 计算技术t的先验概率P(t)。
- 计算技术与攻击之间的关联数量TF(a, t)和TF(obs, a)。
- 计算给定COD时，观察数据obs与攻击a之间的概率P(obs|a)。
- 最终得到技术t在给定COD时的后验概率P(t|COD)。

这种多层朴素贝叶斯推断方法允许利用AttackDB中的因果关系，从而更准确地评估技术与观察数据之间的关联性，进而生成与潜在攻击相关的假设。

方案设计

2.攻击假设生成器：

（2）假设细化

在得到初始假设之后需要对其进行细化。在这个阶段，考虑到技术之间的相互依赖关系，通过增加经常在相同攻击中一起使用的技术的得分来进行假设的细化。为了假设的细化，作者依赖于一组链接预测（LP）技术和相似度度量，包括Jaccard/Tanimoto系数、Adamic Adar、Friends measure/Katz measure和Preferential Attachment。

这些LP技术和相似度度量可以用来预测知识图谱中的缺失关系，或者识别可能不正确的关系。在这种情况下，它们被应用于预测新攻击a_{new}与相关技术之间的关系，并选择排名靠前的技术作为$AT_{\mathrm{a_}new}^{\mathrm{init}}$，从而进行假设的细化。

方案设计

2.攻击假设生成器：

（2）假设细化

作者使用了一系列用于假设生成和细化的技术和算法。如下所示：

Projected Techniques（$\mathrm{proj_T}$）：

这种方法通过将分析师提供的假设或初始假设与每个技术之间的链接权重进行聚合，计算技术的相关性得分。具体地，技术t的pro_{JT}得分由初始假设（AT_{init}）与技术t之间的链接权重的总和计算得出。

Link Prediction on Projected Techniques（$\mathrm{lpproj_T}$）：

该方法在pro_{JT}的基础上利用了技术之间的拓扑结构，通过考虑分析师建议的技术邻域来改进pro_{JT}。通过应用已知的链接预测度量（如Jaccard系数、Adamic Adar、Katz度量和Preferential Attachment）来评估每个技术的可能性。

Link Prediction on Projected Attack（lpproj_A）：

该方法主要是对攻击进行链接预测，以找出与新攻击（a_{new}）最相似的攻击。然后根据这些相似攻击的技术来计算每个技术的得分。

方案设计

2.攻击假设生成器：

（2）假设细化

Projected Attack Techniques（proj_{AT}）：

这种方法根据新攻击与现有攻击之间的相似性，来对技术进行排名。具体地，技术t的ProjAT得分是根据新攻击与包含技术t的攻击之间的Jaccard相似性计算得出。

Supervised Link Prediction（Sup_{LP}）：

这一方法将链接预测问题转化为监督学习问题，并应用随机森林分类器来解决。这种方法可以进一步提高假设生成的准确性。

这些方法结合了不同的技术和算法，以生成和细化关于可能攻击行为的假设。这些假设可以帮助分析人员更好地理解潜在的安全威胁，并采取相应的防御措施。

方案设计

2.攻击假设生成器：

（3）自适应假设细化

初步实验表明，在真实攻击使用大量技术的情况下，细化可能会降低初始假设的质量。因此，引入了自适应细化过程，通过建立一个动态阈值来确定何时进行细化，何时依赖于初始假设。该阈值基于预期与调查攻击相关的技术数量，通过估算每个攻击可能是调查攻击的概率来确定。

TFIDF方法中的假设细化决策：

对于TFIDF方法，决定是否对初始假设进行细化取决于预期的相关技术数量（AT’）和可配置的自适应细化阈值（arth）。只有当AT ‘ ≥ arth时，才执行细化。

方案设计

2.攻击假设生成器：

（4）攻击假设生成

作者通过对监督机器学习(ML)方法的描述来总结攻击假设生成方法。其中CODs被视为实例，而技术被视为标签。每个COD都是一组可观察的信息，特征提取采用了“one-hot encoding”的编码方法，将域名、电子邮件地址、IPv4地址等转换为二进制特征。该研究使用了两种先进的机器学习算法，即XGBoost和Random Forest（RF）来进行假设生成。

04

实验评估

性能评估

评估方法

在实验设置中，首先，作者使用了不同的策略来构建初始假设，这包括：

1. 由模拟人类分析员（H）构建的初始假设（AT_H^{init}）。
2. 通过自动从IoC推断出的初始假设（AT_{IOC}^{init}）。
3. 通过自动从遥测数据推断出的初始假设（AT_{TEL}^{init}）。

对于基于人类分析员的初始假设，假设分析员试图对调查的攻击做出正确的决定。或者犯了两种类型的错误：误报和漏报。这些错误可以通过两个可配置参数来捕获：误报率和漏报率。

针对基于IoC和遥测数据的初始假设推断，我们依赖于前文中介绍的初始假设推理，并使用ML算法（XGBoost和RF）的默认配置。

对当前观察到的日志数据（COD）引入错误。使用不同的误报率和漏报率生成各种输入数据样本。在评估中，使用平均精度作为评估指标来评估TTP推断方法的性能。并计算初始假设和细化假设的AP得分，并计算精炼假设相对于初始假设的AP得分差异，从而评估细化假设相对于分析员假设的改进程度

性能评估

初始假设生成问题结果

图6显示了ioc初始假设生成的性能与观测数据中噪声的关系。

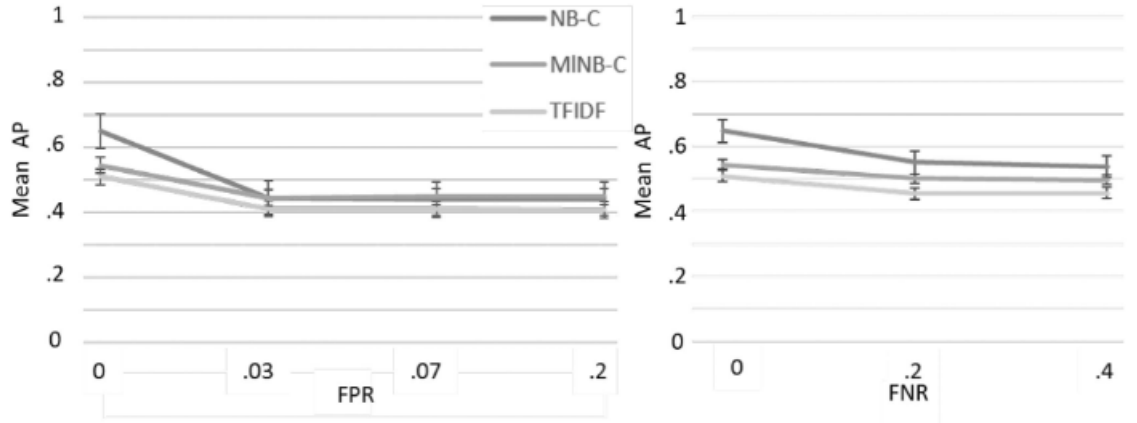


Fig. 6. Initial hypotheses from IoCs. (left) AP as a function of fpr_{COD} when $fnr_{COD} = 0$. (right) AP as a function of fnr_{COD} when $fpr_{COD} = 0$.

性能评估

初始假设生成问题结果

图7给出了每种ih算法的平均AP(从左至右:基于模拟人类分析师(H)的推理，基于IoCs的ih自动推理，基于遥测的初始攻击假设自动推理AT_{init} 和基于机器学习(ML)的ih生成)。

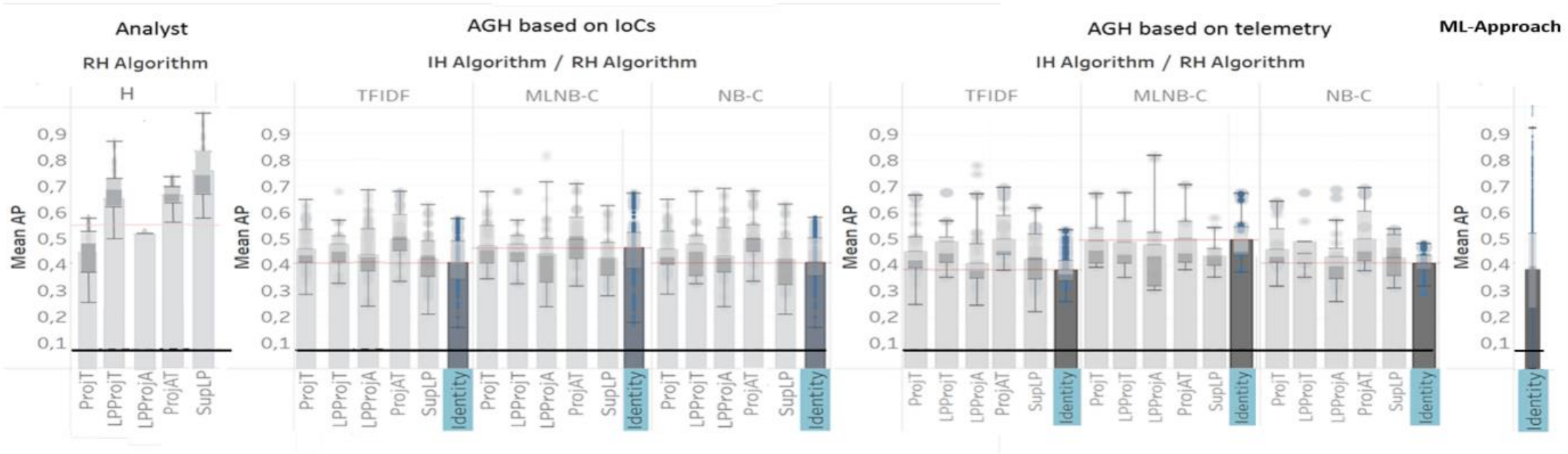
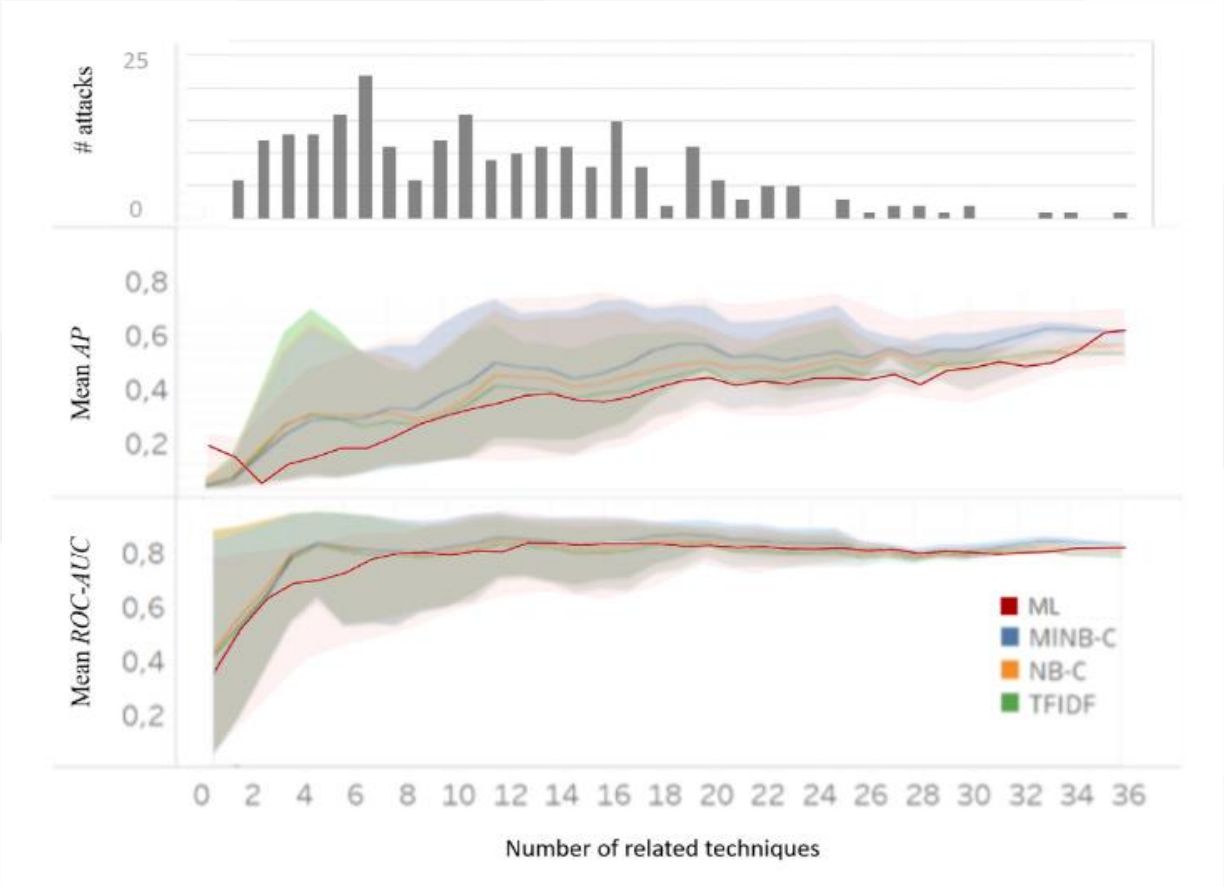


Fig. 7. Mean AP of analyst based (left,) automatic inference of attack hypotheses based on IoCs (middle-left) and Tel (middle-right), and ML based inference of attack hypotheses (right). Black line represents the AP of a random baseline.

性能评估

初始假设生成问题结果

AT初始推断精度与相关技术数量的依赖关系。



05

总结与思考

总结与思考

总结：

本文中作者提出了一个综合的多层次威胁知识库，该知识库来源于多个开源威胁情报源，称为攻击数据库。可以使用AttackDB生成攻击假设，其中包括对所调查攻击的高级描述。作者专注于从攻击系统中发现的可观察工件中推断对抗技术，依赖于多个初始攻击假设生成算法产生的初始假设，之后对初始假设进行细化，最后使用机器学习方法进行最终攻击假设的生成。

思考：

在本文中作者使用威胁情报构建安全专业知识图谱，在图结构的选择是值得我去学习的方向，作者的两部分工作包含了安全知识库和攻击假设生成器，是否可以使用图神经网络或者其他的图算法构建类似于本文中对某些特殊攻击例如APT进行预测与溯源是值得我去思考的。

南京郵電大學
Nanjing University of Posts and Telecommunications

感谢聆听

汇报人：孙艺博