MITRE Engenuity ATT&CK 如何评估 EDR

工作来源

ASIA CCS 2024

工作背景

从 2018 年开始，MITRE 启动了 ATT&CK Enterprise 评估。在评估中，MITRE 重构攻击组织典型的攻击链，对 EDR 产品进行测试。MITRE 官方表示不会对结果进行解读，作者认为这样阻止了用户从结果中直接获取信息。有关该评估的其他相关信息，可以参考以前的文章或者论文原文，此处不再赘述了。

评估存在的主要缺陷是：

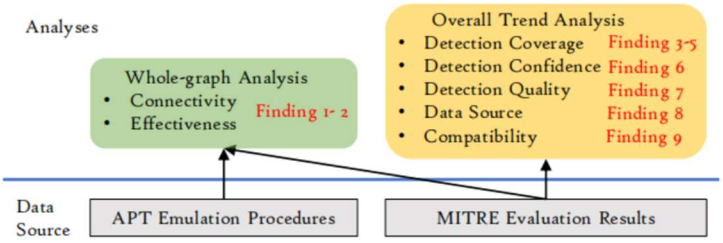
1）缺少全局图分析。单点检测容易被绕过且会带来误报，最先进的 EDR 都将关联图及其衍生模式内嵌在产品中。

2）缺乏完整的解释。没有给出官方解释说明导致各家公司都表示自己取得了近乎完美的表现，各方都选取了特定的表述角度宣传自身产品。

3）评估表述不一致。几乎每次评估的方法和表述中使用的术语定义都在变化，导致结果之间存在差异。

工作设计

整体分析如下所示。



全局图分析

首先按照控制流和数据流的视角来构建图，如下所示。紧接着判断 EDR 是否能完整重建攻击链，以及能否有效聚合攻击链判断攻击严重性。

总体趋势分析

从检测覆盖率（可见性与分析覆盖率）、检测置信度（置信度越高，提供的细节越多）、检测质量（检测延迟低且开箱即用，表示质量较高）、数据源和兼容性上进行总体分析。

工作准备

所有数据的基本情况如下所示：

Round of Evaluation	Participants	Steps	Techniques	# of Detection Made
APT3 (2018)	12	136	51	1970
APT29 (2019)	21	134	53	3982
Carbanak+FIN7 (2020)	29	174	46	7350
Wizard Spider+Sandworm (2022)	30	109	46	3098
Total	37	N/A	82	16.4k

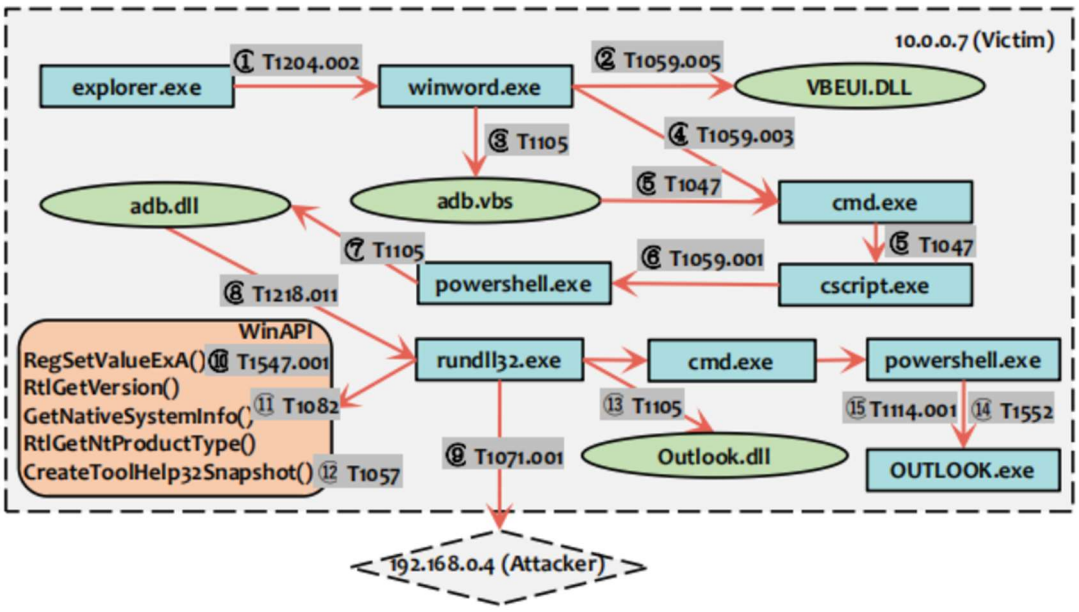
工作评估

全局图分析

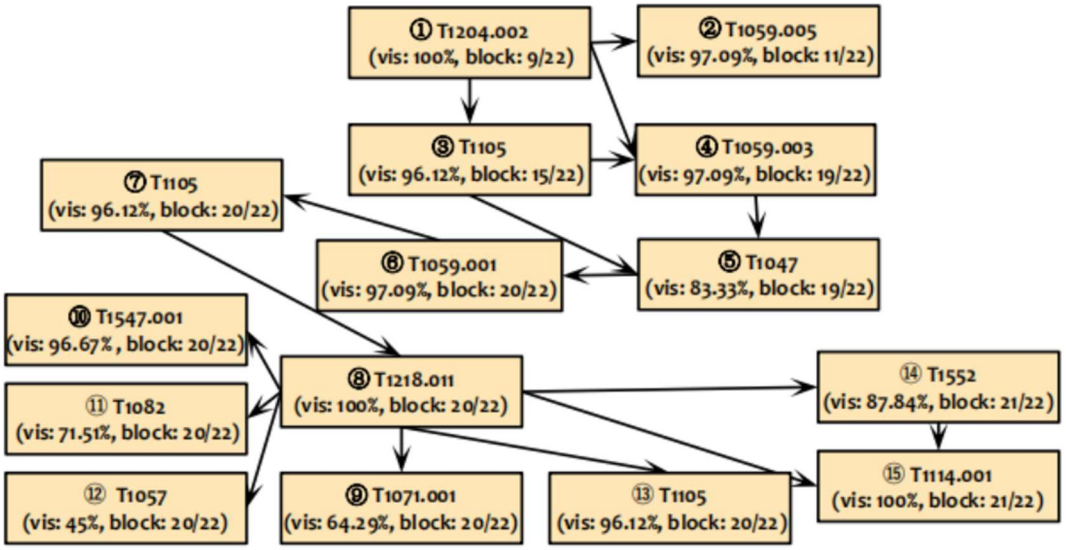
以 2022 年 Wizard Spider+Sandworm 评估为例，环境里的六台主机，有一台是 Linux 的。30 个参评厂商，只有 22 个支持 Linux 数据收集检测。结果是一共 25 个（83.3%）厂商能够获得包含所有攻击步骤的连通图，说明大部分厂商还是可以看到攻击步骤间的联系的。

如下图所示的场景：攻击者通过电子邮件发送了 Word 文档附件，文件中包含经过混淆的 VBA 宏代码。用户下载执行后，会下载 Emotet 恶意 DLL。恶意软件会修改注册表、获取进程信息，还会下载其他恶意 DLL 文件检索 Outlook 凭据。

22 个 EDR 中有 21 个阻止了攻击，大多数都在 Explorer 下载 Word 文档时就能触发拦截了。Word 文档下载恶意 DLL 文件与 VBS 文件时，有 19 个 EDR 进行了拦截。



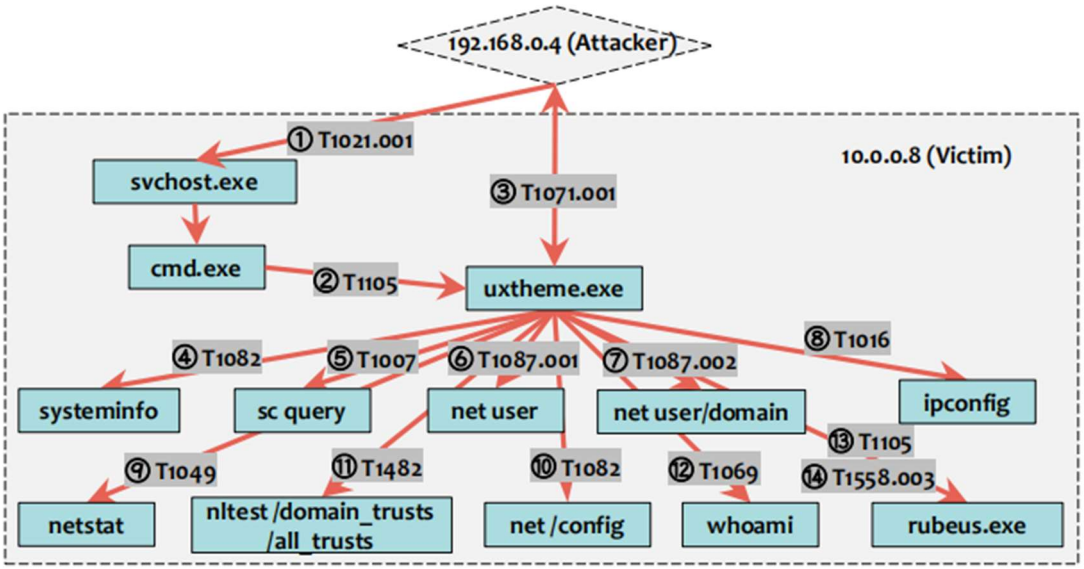
(a)



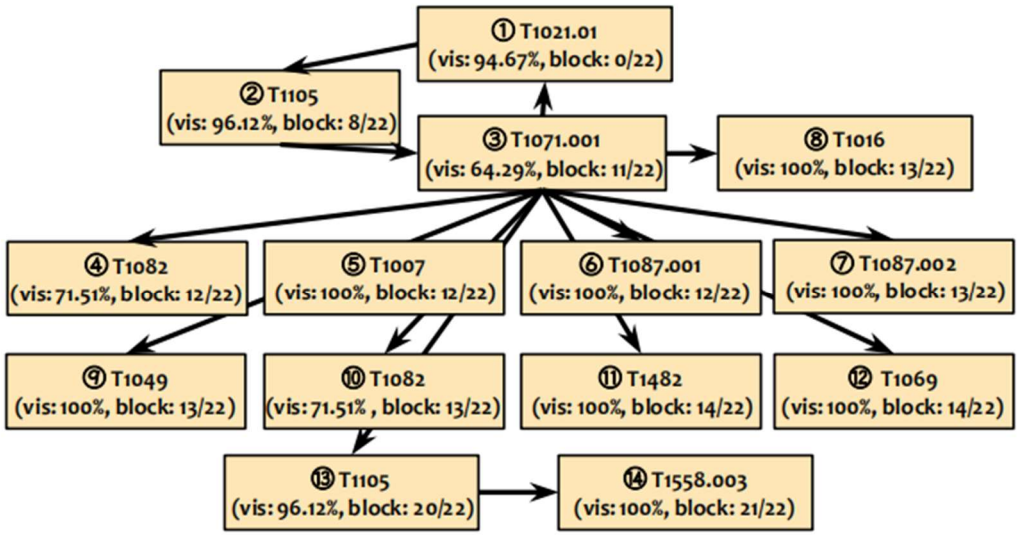
(b)

如下图所示的场景：攻击者利用窃取的凭据入侵受害者主机，随后下载并执行 TrickBot。恶意软件收集各种信息后，下载名为 Rubeus 的工具来窃取加密凭据。

22 个 EDR 中有 21 个阻止了攻击。只有一半的 EDR 能够在恶意软件回连时进行拦截，剩下一半基本都在收集各种信息时进行拦截。



(a)



(b)

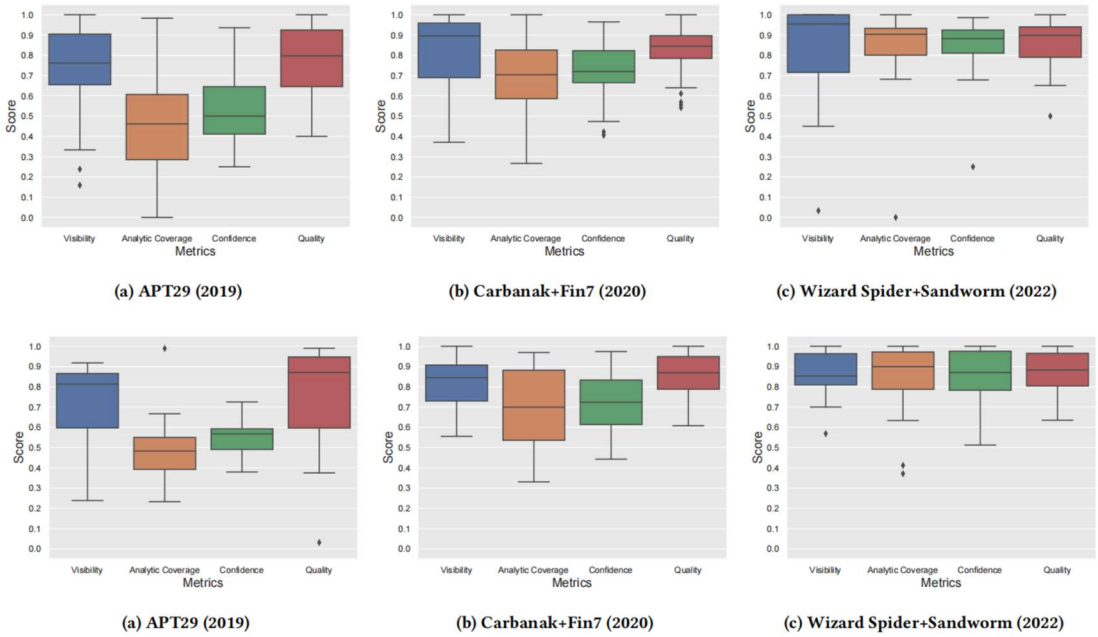
第二个场景下的检测延迟明显要高，因为第二个场景中第三步的可见性只有 64.29%，可见性降低导致 EDR 没能收集足够的信息进行检测。

测试 3、4、5 的保护率明显下降，例如测试四中只包含两个步骤：转储整个 C 盘和注册表，可能 EDR 认为这尽管可疑但不足触发告警。在下载执行恶意文件时经常触发拦截，但这样应对无文件很可能存在问题。

Test	# of Blockage	# of Participants	Protection Rate
1	21	22	95.5%
2	21	22	95.5%
3	16	22	72.7%
4	12	22	54.5%
5	15	22	68.2%
6	20	22	90.9%
7	9	17	52.9%
8	20	22	90.9%
9	18	22	81.8%

EDR 需要攻击者在系统内表现出更长的攻击链，才能收集到足够多的信息来进行拦截。这样可能会导致即使他们发现了恶意行为，但并不会进行拦截。

总体趋势分析



(a) APT29 (2019) (b) Carbanak+Fin7 (2020) (c) Wizard Spider+Sandworm (2022)

(a) APT29 (2019) (b) Carbanak+Fin7 (2020) (c) Wizard Spider+Sandworm (2022)

① 可见性

从技术角度来看，可见性中位数为 95% 意味着 95% 的 EDR 系统能够看到一半的攻击步骤。从厂商角度来看，可见性中位数为 85% 意味着一半的 EDR 能够看到 85% 的攻击步骤。

Carbanak+FIN7 评估时，只有 40% 的 EDR 能够记录加密信道的 C&C 通信，而后来大概 80% 的 EDR 都对各种可能的协议进行了记录与检测。可能目前 EDR 系统仍是有选择地收集传输层的流量，而忽略应用层协议。

Wizard Spider+Sandworm 评估时，15 种攻击技术实现了 100% 的覆盖。所有的 EDR 系统都重视进程的加载与执行，与执行相关的技术都是可见性最好的。

从结果上来看，AhnLab 几乎不监控文件系统，与收集凭据相关的大多数技术都不可见。AhnLab 还会对某些 PowerShell 的下载告警，但相同 IP 地址的其他 PowerShell 下载不告警。不只是 AhnLab 将可见性从 0.517 提升至 0.761，全行业的可见性都在提升。

某些技术(如 Archive Collected Data)在 Carbanak+FIN7 时获得了 100% 的可见性，但在 Wizard Spider+Sandworm 时可见性跌到 0.033%，这表明 30 个 EDR 中只有一个是可见的。这种巨大的差异表明，相同的技术有许多不同的实现方式，一旦更换可能对检测能力就是巨大的挑战。

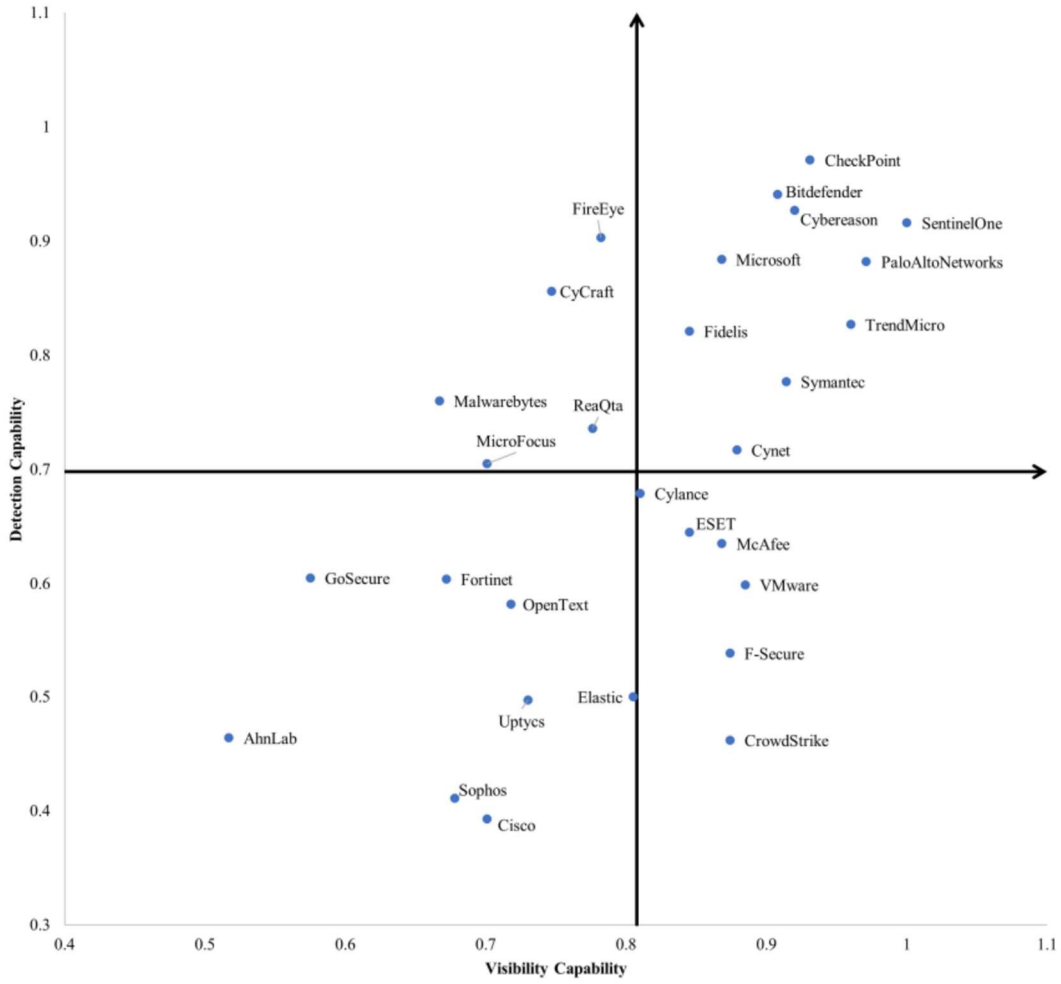
② 分析覆盖率

这几年的评估表明分析覆盖率大幅提升，平均值从 69.8% 上升到 85.5%。超过 90% 的 EDR 会对 50% 的可见攻击步骤进行告警，50% 的 EDR 会对 90% 的可见攻击步骤进行告警。

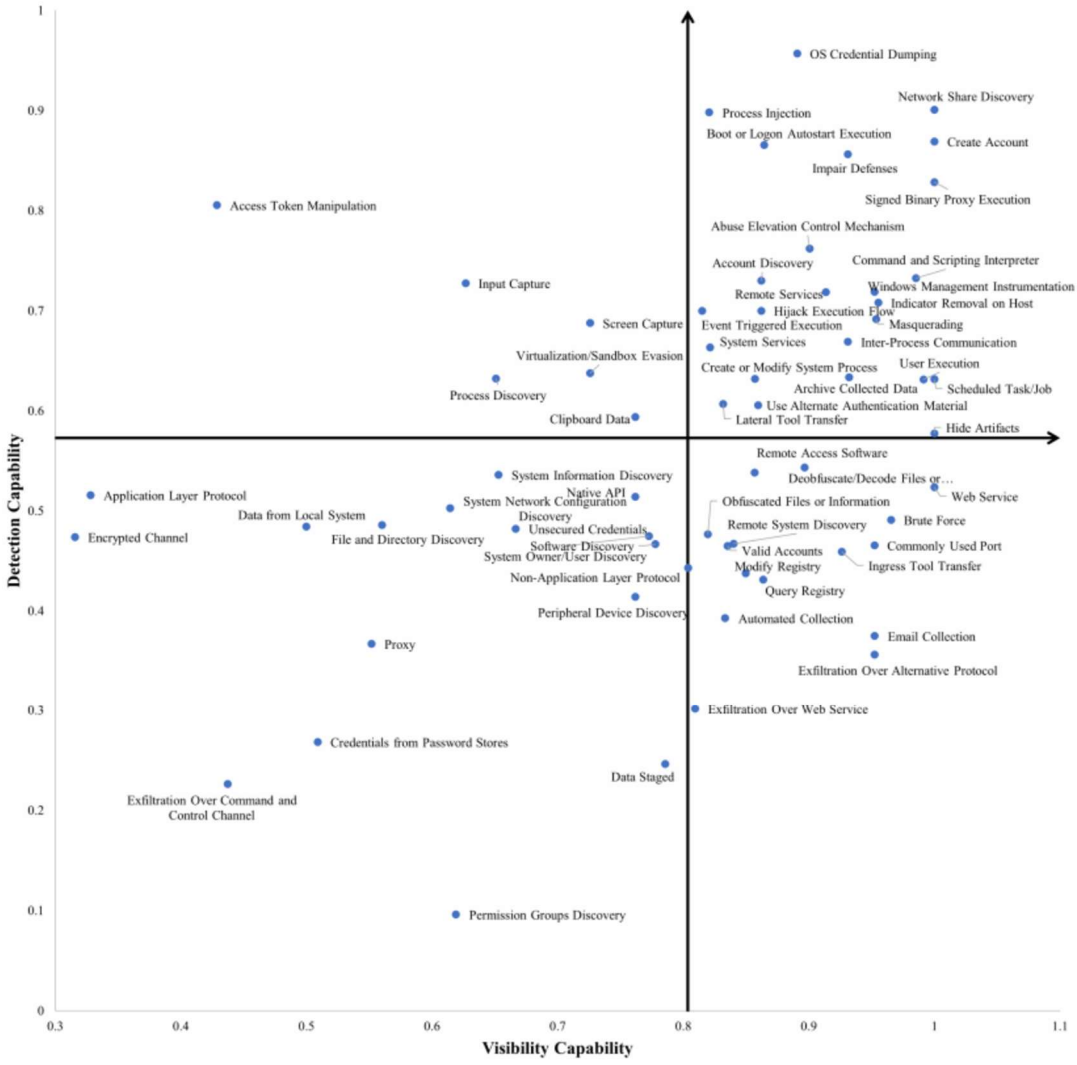
分析覆盖率较低的都是后渗透阶段，如 C&C、数据外带等。分析覆盖率较高的都是攻击链的前部分，如防御规避、凭据访问等。这说明在后渗透阶段的攻击行为与正常行为很难区分开，确实很难检测。

各厂商的象限图如下所示。SentinelOne 与 PaloAlto Networks 等厂商都是领

跑第一象限的，第二象限的典型代表 FireEye 分析覆盖率高于平均水平但可见性落后，第四象限的典型代表 CrowdStrike 可见性高于平均水平但分析覆盖率低。



各技术的象限图如下所示。63 项技术中有 42 项技术都在第一、三象限中，可疑近似认为可见性与分析覆盖率是强相关的。6 项技术位于第二象限，意味着 EDR 很难收集相关数据，一旦可见会立刻判黑。15 项技术位于第四象限，意味着 EDR 尽管看到了这些行为，但难以判断是不是恶意的。



③ 置信度

遥测对应 25% 置信度，行为对应 50% 置信度，战术对应 75% 置信度，技术对应 100% 置信度。

与操作系统凭据转存和网络共享发现相关的恶意行为 EDR 基本都能识别，并且能够通过其他信息补充告警。但与密码凭据转存和权限发现相关的恶意行为则难以被检测，更不用说补充其他信息。

Palo Alto Networks、Sentinel One 和 CheckPoint 等优秀的 EDR，在置信度和分析覆盖率上表现都很好，不仅能够检出恶意行为还能提供丰富的细节。置信度从 50% 提升到 88.25%，这意味着检测细节的水平显著提高。最初 EDR

只能在行为以上级别检测到 50% 的攻击步骤，后来 EDR 能够在技术级别检测到 75% 的攻击步骤。

④ 检测质量

整体来看，检测质量还是一直在提升的。检测质量与自适应能力有关，检测质量越高越可以减少人工干预，否则需要大量的人工调整和分析。

⑤ 数据源

2019 年评估 APT 29 时有 9 个数据源，2022 年已经增加到 42 个数据源。不仅数量在增长，粒度也更细加了，当然进程、文件、网络、系统调用/API 仍然是 EDR 最有价值的数据源。

⑥ 兼容性

在测评中针对 Linux 平台的攻击也与针对 Windows 平台的攻击很相似，但在 Windows 上能防住的到了 Linux 上有一半都防不住。Linux 平台上数据收集与系统防护的能力还需要改进，目前的效果是比较差的。

工作思考

全局图关联能力对防御来说至关重要，单个步骤无法为拦截响应提供足够的信息。即便 EDR 有全局图关联能力，但也存在检测延迟、无法拦截以及不能跨主机关联的问题。

EDR 的数据收集能力在逐年提升，但同一技术的多种实现方式可能覆盖尚不完全，检测粒度需要进一步细化。

EDR 对检测典型的恶意行为很有信心，但是如果恶意行为与良性行为不那么明显，检测是很含糊的。