执行体全量识别与精细管控

执行体的识别与管控

——IT治理的基石

安天 | 技术委员会

引子——为什么我们要关注执行体

- 网络安全对抗的本质，在过去和未来非常长的一个阶段，都是代码对抗。
- 执行体既是代码对抗中的攻击目标，也是“武器化”攻击装备，同时也防御机制的承载者。
- IT治理的关键抓手，包括暴露面、脆弱性、补丁、漏洞等，都与执行体相关。
- 但与此同时，我们对执行体的认识还是高度不全面、不完备，缺少系统方法的。

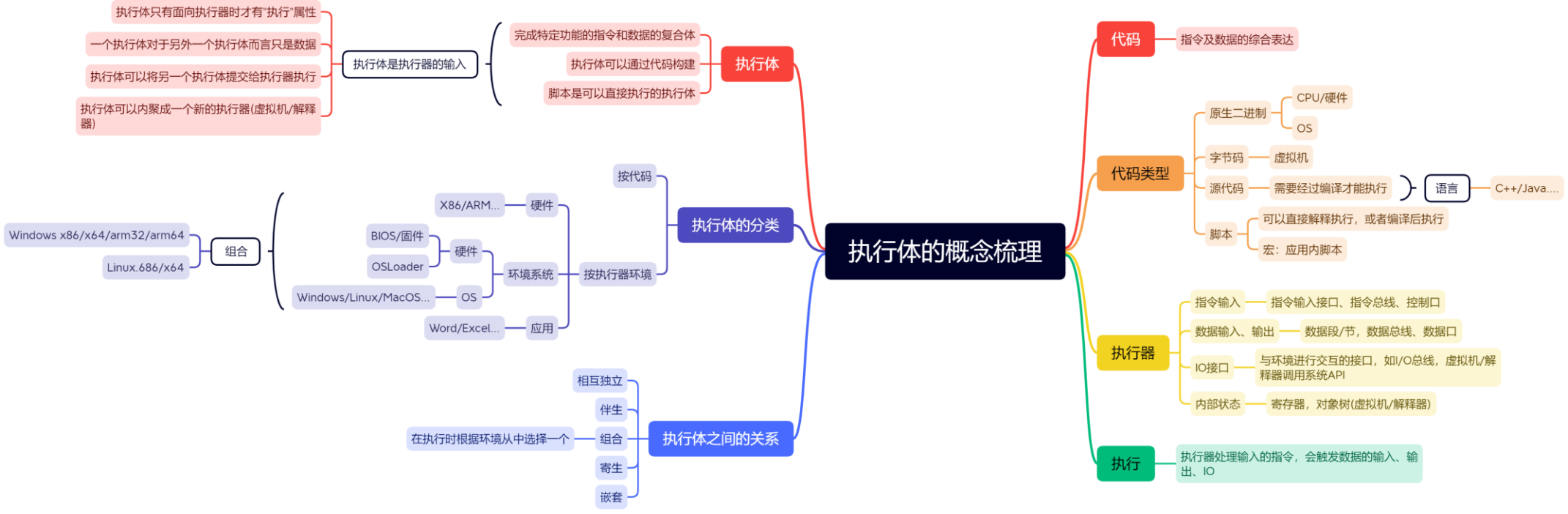
目 录

01 执行体的概念、多各角度的观察与执行体治理

执行体相关的定义和概念

- 执行体是为**实现特定目的代码和数据的综合表达**，由硬件系统、固件系统、操作系统、应用程序或虚拟机等执行环境执行。
- 执行体可以**在执行环境中独立执行**，也可以**嵌入在其他执行体或数据中执行。**
- 执行体在网络中以**文件形态和非文件的嵌入式形态**存在。
- 各种数字化系统（信息系统）往往由**一个或多个执行体构成。**
- 攻击可以**通过恶意代码执行体实现**，也可以**借助正常“执行体”实现。**
- 执行体的对象形态是**复杂多维**的，执行体的行为是**动态变化**的。

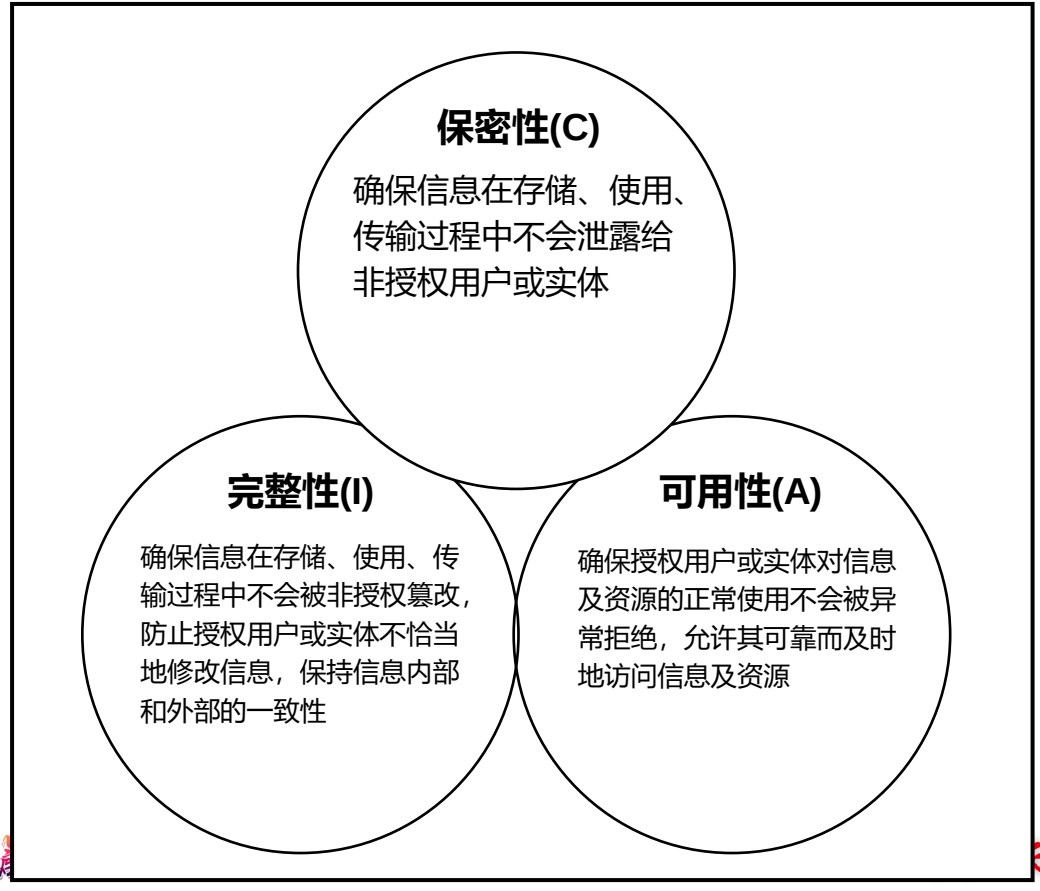
概念本源视角：从运行机理和产生过程看执行体



度量要素视角：信息的安全要素和执行体安全的要素的对比

信息的安全三要素（形成与70年代，并逐步演进为共识）

信息是可度量的



执行体安全的四要素（安天于2023年1月抛砖引玉）

执行体是不可完全度量

恶意性(M)

以恶意为目的编写，会产生不同程度的危害影响。

脆弱性（V）

执行体缺陷缺陷和漏洞，可能被威胁利用对资产造成损害

风险性(R)

但运行结果可能对对环境和资产带来一定的影响。

可信性(D)

是指生产、加工和分发、使用的过程，导致产生不同的可信任程度。

攻击视角：杀伤链是一个构造执行体作用于执行体组成的系统的过程

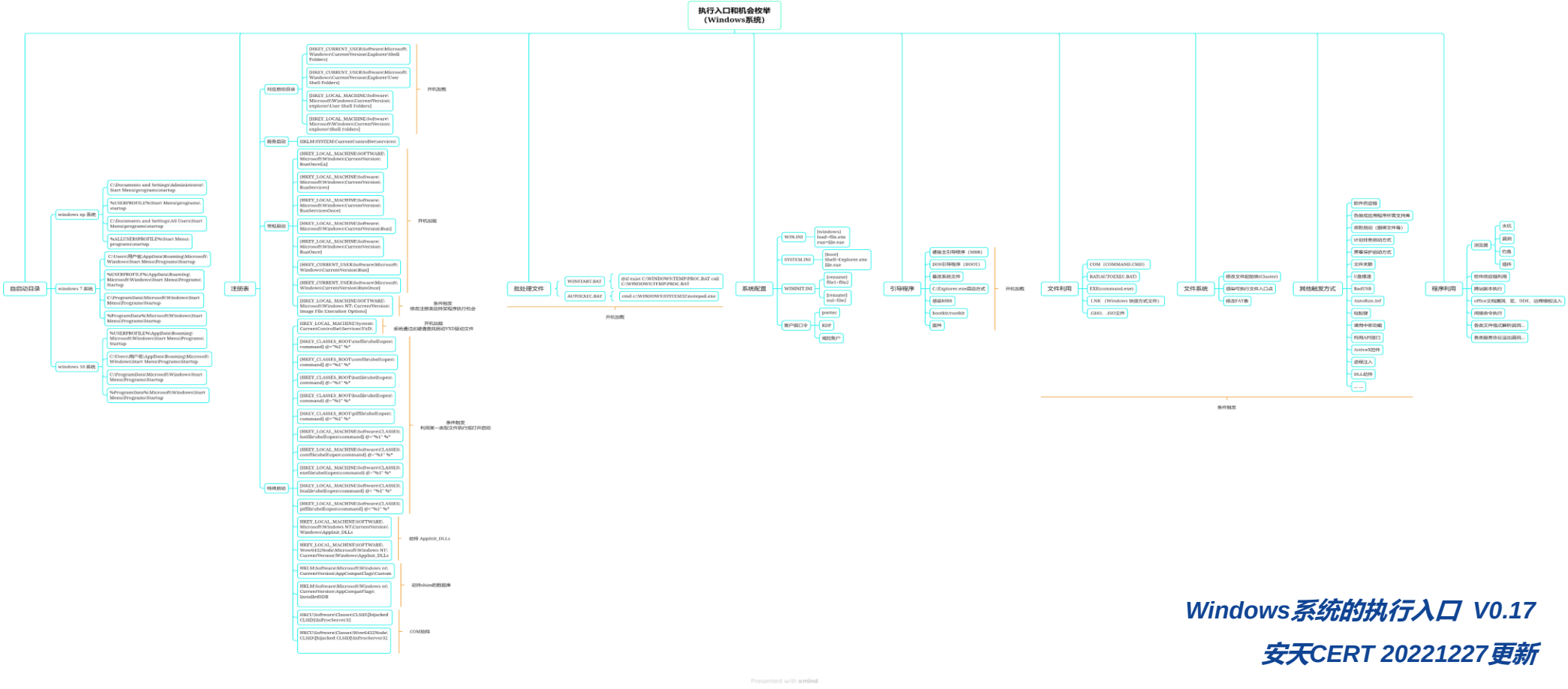
	侦察跟踪	武器构建	载荷投递	漏洞利用	安装植入	命令控制	致效应用
攻击武器	主动扫描扫描器	攻击平台 恶意代码 漏洞利用工具	恶意代码（传输中）	Shellcode	远控木马	恶意代码的信道构建	基于指令和预设逻辑达成预设效用
辅助工具	SBOM工具	开发编译器库工具	捆绑器 加壳器	Exploit生成工具	打包软件、宿主环境、解释器	反弹Shell生成器	被利用的其他合法软件
用户资产	系统运行程序的版本指纹等遭获取	用户系统和应用环境	浏览器、邮件客户端、提供公共访问服务的程序	系统程序/网络服务/文件数据等	系统执行入口 应用环境入口 可持久化位置	失陷业务系统、跳板	系统管理工具、提供数据存放能力的服务……

攻击视角：绝大部分网络攻击技战术动作依赖于执行体完成

形态视角：执行体的对象形态

执行体形态与模式	存在形式	格式	执行环境/平台	加载来源	例子	相关攻击组织	攻击事件
Windows可执行文件	文件	二进制	操作系统执行	磁盘	cmd.exe	绝大多数APT组织	白象的舞步使用大量开源和自研木马发起攻击活动
Windows驱动程序文件	文件	二进制	操作系统执行	磁盘	win32k.sys	极少数高级APT组织	震网病毒攻击伊朗核工业控制系统使用驱动进行持久化和反检测
脚本文件	文件	文本	应用程序解释执行	磁盘	slmgr.vbs	绝大多数APT组织	污水/MuddyWater针对中东的系列攻击活动
批处理文件	文件	文本	应用程序解释执行	磁盘	msdtcvtr.bat	部分APT组织使用	脑肚虫组织针对军事人员攻击活动
AutoCAD插件	文件	文本	应用程序解释执行	磁盘	acad.lsp	绝大多数APT组织	苦象组织使用带有恶意宏的文档针对南亚国家攻击
宏	非文件/嵌入式	文本或二进制	应用程序解释执行	磁盘	xlsm文件中的宏	绝大多数APT组织	苦象组织使用带有恶意宏的文档针对南亚国家攻击
引导记录	非文件/嵌入式	二进制	固件系统执行	磁盘	MBR	部分APT组织使用	拉撒路组织使用MBR擦除攻击韩国政企机构
UEFI驱动	文件	二进制	固件系统执行	磁盘	bootx64.efi	极少数高级APT组织	Hacking Team泄漏的“MosaicRegressor”
UEFI程序库	文件	二进制	其他执行体调用	磁盘	kdstub.dll	极少数高级APT组织	Hacking Team泄漏的“MosaicRegressor”
微码	非文件/嵌入式	二进制	固件系统执行	固件	.bin CPU微码文件	极少数高级APT组织	2017年“34C3: Hacking Into A CPU’s Microcode”
PXE引导程序	非文件/嵌入式	二进制	固件系统执行	网络	pxelinux.0	极少数高级APT组织	Vbootkit工具
APK程序	文件	复合	混合执行	磁盘	com.ss.android.ugc.aweme_23.9.0_230901.apk	部分APT组织	Kimsuky组织使用移动端恶意代码针对韩国东亚研究所的恶意活动
JAR程序	文件	复合	虚拟机执行	磁盘	start.jar	部分APT组织	海渊使用Python伪编译的EXE文件攻击工控PLC系统
小程序	文件	文本	混合执行	网络	微信小程序	部分攻击组织	某组织利用微信小程序传播赌博技巧引诱用户进行博彩
WASM程序	文件	二进制	虚拟机执行	网络	fivenightsatplant.wasm	部分攻击组织	某组织使用WASM编写键盘记录器、加密货币挖矿程序
容器镜像	文件	复合	其他执行体调用	磁盘	Gitlab15.4.tar	部分APT组织	TeamTNT组织使用包含而恶意代码的Docker镜像进行活动
SQL代码	非文件/嵌入式	文本	应用程序解释执行	内存	各种ORM生成的SQL语句	部分APT组织	方程式组织攻击EastNets事件
SQL注入	非文件/嵌入式	文本	应用程序解释执行	网络	1' union select 1,database()#	绝大多数APT组织	APT28、APT39在历史攻击活动中使用SQL注入攻击对目标网站进行攻击
DDE漏洞利用代码	非文件/嵌入式	文本	应用程序解释执行	内存	DDEAUTO c:\\windows\\system32\\cmd.exe	绝大多数APT组织	白象、拉撒路等多个组织使用过DDE漏洞发起攻击活动
安装包	文件	复合	其他执行体调用	磁盘	setup.msi	部分APT组织	蔓灵花APT组织针对国内的攻击活动使用过MSI格式木马
伪编译可执行文件	文件	复合	混合执行	磁盘	Test.pyz	部分APT组织	海渊使用Python伪编译的EXE文件攻击工控PLC系统
SNAP应用	文件	复合	操作系统执行	磁盘	Ubuntu Snap Store	部分攻击组织	Ubuntu Snap 应用商店上发现加密矿工事件
eBPF程序	非文件/嵌入式	二进制	虚拟机执行	内存	BPFDoor、Symbiote	极少数组织	利用eBPF运行木马后门
裸机程序	非文件/嵌入式	二进制	固件系统执行	固件	无	极少数高级APT组织	震网病毒篡改MC7格式代码攻击伊朗核工业系统PLC
BadUSB	非文件/嵌入式	复合	固件系统执行	固件	橡皮鸭	极少数高级APT组织	NSA水蝮蛇USB攻击设备
一句话木马	非文件/嵌入式	文本	应用程序解释执行	网络	<?php @eval($_POST['cmd']); ?>	绝大多数APT组织	Kimsuky组织针对韩国新闻行业的钓鱼活动
网络服务溢出	非文件/嵌入式	二进制	其他执行体调用	网络	EDB-ID-50944 qdPM 9.1 RCE	绝大多数APT组织	永恒之蓝漏洞利用
文件格式溢出	非文件/嵌入式	二进制	其他执行体调用	磁盘	OFFICE、PDF、SWF、LNK格式文件	绝大多数APT组织	绿斑组织针对我国的鱼叉钓鱼活动
二进制ShellCode	非文件/嵌入式	二进制	其他执行体调用	内存	EDB-ID-49855 Linux/x86 ShellCode	绝大多数APT组织	FIN7组织使用XLL文件加载Shellcode代码的攻击活动

执行机会视角：现代操作系统为执行体提供了丰富的入口机会



Windows系统的执行入口 V0.17

安天CERT 20221227更新

什么是执行体治理

- 执行体治理是**网络安全运营者通过识别和管控执行体保障网络安全的持续过程。**
- 持续过程不仅仅要完成检测、防御、清除恶意执行体和控制非恶意执行体网络访问等**基础防护**，更要建立识别、塑造、检测、防御和响应的**流程闭环**。在流程运行闭环的基础上，全面掌握执行体的静态分布情况与业务应用的执行体构成，**建立信誉清单**，同时能够**识别执行体的执行动作并依据基线进行控制。**
- 在基线建立之后，**识别全部执行体**，全面掌握执行体和执行体的行为与业务之间的**支撑关系**，建立信誉指标、行为指标、业务影响指标等**量化指标**，以指标为指引针对不同场景**建立配套的管控规则库、基线库和模型库**，并**持续运营**实现能力与时俱进、效能不断提升。

应用与执行体是关键治理层次、执行体是关键治理对象

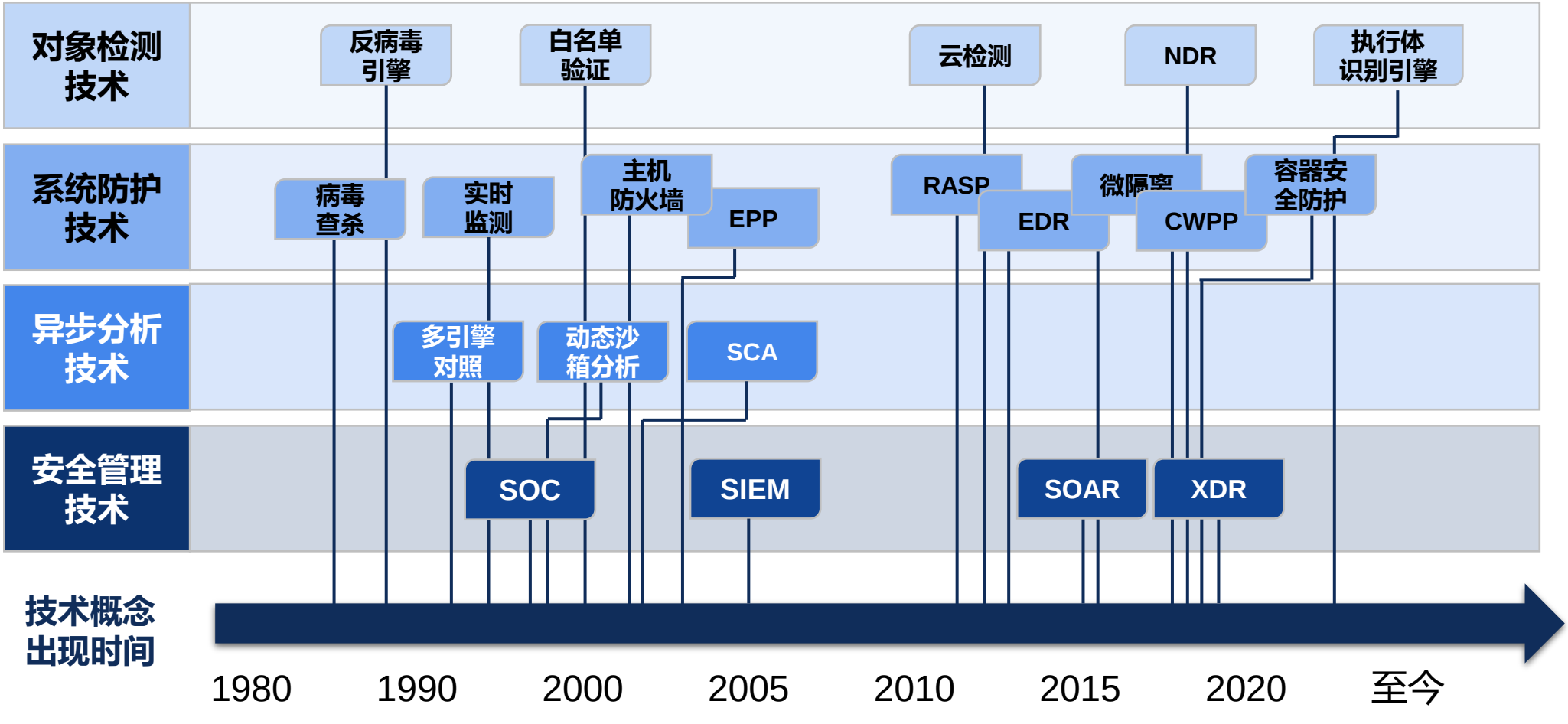
安全治理		安全价值				
维度	举措	业务与数据	身份与凭证	网络与地形	应用与执行体	资产与系统
业务与数据	•WAF •脆弱性扫描 •数据库审计 •数据分级分类 •安全配置加固 ……	/	•业务应用的用户身份识别 •数据的用户身份主体管理 ……	•拒止业务应用的非法网络访问 •约束应用数据的网络通联范围 ……	•应用与执行体脆弱性修复 ……	•清晰化资产的业务子地形 •清晰化资产的数据地图与分布 ……
身份与凭证	•用户与身份管理 •证书管理 •访问凭证管理 ……	•业务与数据的用户身份管理 •限制业务与数据访问权限 ……	/	•网络用户的准入管控 •域用户网络访问权限管控 ……	•应用系统用户与身份管控 •应用与执行体可信性管控 ……	•管控基础设施的身份与凭证 •管控操作系统的身份与凭证 •管控域用户的身份与凭证 ……
网络与地形	•网络防火墙 •主机防火墙 •流量检测与响应 ……	•约束业务与数据网络通联 •清晰化业务应用的访问关系 •检测与拒止网络攻击 ……	•用户访问关系清晰化 •异常用户网络访问检测 ……	/	•应用与执行体网络地形 •管控应用与执行体网络访问……	•清晰化资产的网络地形 •管控资产网络访问 ……
应用与执行体	•反病毒引擎、EDR •文件沙箱 •软件成分分析 •企业应用商店 ……	•清晰化业务应用软件构成 •细粒度管控业务运应用的可执行范围 •管控数据访问行为 ……	•防护场景的数字签名验证 ……	•管控应用与执行体网络访问 •清晰化应用与执行体网络地形 ……	/	•清晰化资产系统的全量执行体 ……
资产与系统	•IT资产管理系统 •漏洞与补丁管理 ……	•清晰化资产的业务子地形 •清晰化资产的数据地图与分布 ……	•提升资产用户身份可见性 •降低身份仿冒风险 ……	•收缩资产与系统网络暴露面 ……	•清晰化应用与执行体 •应用与执行体漏洞修复 ……	/

执行体治理对安全能力的价值意义

	识别	塑造	防护	检测	响应
执行体治理对各个环节的收获	• 有效澄清系统环境、业务环境：掌握执行体的行为与业务之间的支撑关系；理解执行体及其所需权限； • 执行体具备的能力及其与脆弱性、暴露面的对应关系等。	• 连接管理：依据目的管控连接。 • 收敛暴露面：管控执行体开放服务及存在执行更新能力的通道。 • 减少信息外溢：识别访问隐私数据行为，管控传输用户隐私动作。	• 行为判断：识别执行体的资源访问、连接、创建、写入、执行的客体，判断其行为目的，对违规行为进行拒止。 • 预判拒止动作效果。	• 基于执行体分布和行为监测，筛选出值得关注的、未知的执行体。 • 提供异常行为依据。 • 基于执行体行为和用户操作，推断用户行为及其目的。	• 基于执行体的潜在和激活能力、创建信道等手段，支撑追溯攻击来源。 • 基于攻击执行体的行为和潜在行为，支撑环境和数据恢复、策略调整。

02 执行体治理的历史演进过程与现实复杂性

支撑技术能力演进路径


对象检测技术
反病毒引擎
白名单验证
云检测
NDR
执行体识别引擎
系统防护技术
病毒查杀
实时监测
主机防火墙
EPP
RASP
EDR
微隔离
CWPP
容器安全防护
异步分析技术
多引擎对照
动态沙箱分析
SCA
安全管理技术
SOC
SIEM
SOAR
XDR
技术概念出现时间
1980
1990
2000
2005
2010
2015
2020
至今

执行体治理相关技术和模式的作用价值和局限性

时间	价值/不可替代性	盲区与局限性
反病毒引擎技术	• 对海量已知威胁的精准识别（分类、家族、变种） • 执行体的历史应用和攻击活动关联（配套知识） • 技战术揭示和攻击组织的精准指向（配套知识）	• 可免杀绕过的资源 • 对I/O的依赖
威胁情报技术	• 具有明确的信标指向性 • 便于快速运营和消费	• HASH规则的鲁棒性奇差 • Tools情报缺少统一的机制 • TTPS层面情报难以消费
系统主防技术	• 执行体动作级的发现和评价，可以发现执行体的资源访问和对系统的操作等 • 防御环境的主动塑造 • 攻击杀伤的实时拒止	• 常态资源占用 • 对业务连续性产生潜在影响
可信计算技术	• 执行体发布者身份的鉴别 • 执行体启动时和运行时的可信验证	• 目前缺少对文本、宏等的有效验证机制 • 证书盗用、证书滥用等问题难以 • 一旦可信根出现问题，将影响整个信任链，可能对整个系统产生安全风险
内存对抗技术	• 降低缓冲区溢出攻击的成功率	• 对应用漏洞无效 • 对独立文件载体的恶意代码执行无效

执行体治理模式的演进

	主机杀毒阶段	综合主机防御阶段	零信任与全量执行体治理
场景特点	个人计算革命，以DOS系为主机操作系统，有限的局域网络能力。	信息高速公路建设，Windows、Linux等系统成为主流，主机系统成为为互联网节点。	万物互联带来海量流量，云应用及大数据应用得到极大进展
威胁状况	恶作剧炫技攻击为主。 数万种变种恶意代码。	追求经济利益。数百万种变种恶意代码。 国家背景的定向APT攻击。 百万量级恶意代码	复杂的统计模式全面合流。 数千万种变种恶意代码。
端点安全环节	AV	EPP、EDR、CWPP	UES、统一工作负载防护、微隔离
执行体治理模式	黑名单（病毒规则）	签名验证、安全基线和主防	全量执行体识别和管控治理

场景环境复杂度

治理难的原因	具体情况	场景案例
规模庞大	不断产生新的增量	政企场景网内评价每月每端点约新增超200个执行体
	存量巨大	单端点级别的IT场景下网内各类执行体消重后总量达数十万量级
	同类的不同版本	某软件同时存在数个版本（漏洞、篡改风险）在不同终端运行
形态与来源复杂	复杂的来源	采购安装、网管安装、用户自行安装、用户自研、攻击者投放、捆绑劫持等非客户意愿安装
	复杂的形态与格式	网内存在各类执行体形态数十种、文件格式上百种
完整验证困难	签名滥用	存在具有签名的恶意、灰色程序（管理不规范的证书颁发）
	签名盗用	使用盗用签名的恶意程序（震网mrxnet.sys带有Realtek签名）
	无签名	大量黑白未知的无签名文件
	可签名但未签名文件	微软、谷歌等软件厂商的正常但无签名文件
	签名验证机制利用	Flame恶意软件利用微软签名验证漏洞绕过验证
使用、运行环境复杂	有风险但无法禁止的文件	云同步软件、输入法、旧版存在漏洞单无法更新的软件
	信誉的场景个性化	非系统路径出现系统程序或第三方程序

场景复杂性x规模是本质挑战

操作系统：Windows 10 企业版 21H2　　系统安装时间：2022/3/1　持续监测时间：5天

统计	执行体总数	可执行程序	模块	系统文件	驱动文件	工具软件	办公软件	业务软件	文档	脚本	其他应用
综合统计	12W	1.5W	9.5W	5.6W	480个	1800个	9000个	8000个	1800	5600个	2.3W

统计	有签名	无签名	有效签名	无效签名	过期签名	吊销签名	证书格式异常	风险签名	主流厂商签名	生僻签名	配置清单中的签名
数字签名统计	7.3W	4.7W	6.2W	2600个	300个	15个	8个	1个	5.8W	120个	1300个

统计	编译器种类	壳种类	供应商	共享文件	隐藏文件	正版软件	未授权软件	持续活跃	偶尔活跃	持续静默	静默
按属性统计	36个	124个	92个	2500个	3200个	92个	4个	6000个	3.4W	3W	5W

统计	有已知漏洞	敏感命令创建的文件	服务执行	启动项执行	生僻技术栈	生僻编译器	生僻壳	生僻厂商	名称伪装	图标伪装	敏感端口监听
安全性统计	7个	600	370个	136个	3个	4个	2个	4个	8个	73个	3个

典型办公主机

操作系统：Windows 11 家庭版 22H2　　系统安装时间：2020/1/15　持续监测时间：5天

统计	执行体总数	可执行程序	模块	系统文件	驱动文件	工具软件	办公软件	业务软件	文档	脚本	其他应用
综合统计	26W	3W	23W	5.8W	640个	4.7W	2000个	300个	2300	7700个	1.2W

统计	有签名	无签名	有效签名	无效签名	过期签名	吊销签名	证书格式异常	风险签名	主流厂商签名	生僻签名	配置清单中的签名
数字签名统计	20.3W	5.7W	18.2W	6000个	230个	37个	9个	3个	9.8W	2000个	230个

统计	编译器种类	壳种类	供应商	共享文件	隐藏文件	正版授权	未授权软件	持续活跃	偶尔活跃	持续静默	静默
按属性统计	55个	114个	125个	1.3W	1200个	67个	37个	8000个	11.4W	6W	13W

统计	有已知漏洞	敏感命令创建的文件	服务执行	启动项执行	生僻技术栈	生僻编译器	生僻壳	生僻厂商	名称伪装	图标伪装	敏感端口监听
安全性统计	34个	33个	245个	138个	9个	3个	6个	19个	19个	3个	4

居家办公个人主机

操作系统：统信服务器操作系统V20 1050天　　系统安装时间：2021/01/15　持续监测时间：5

统计	执行体总数	可执行程序	模块	系统文件	驱动文件	工具软件	办公软件	业务软件	文档	脚本	其他应用
综合统计	8W	1.1W	6.9W	3.2W	320个	6850个	356个	8650个	1200	1.2W	4000个

统计	有签名	无签名
数字签名统计	5.8W	2.2W

由于统信系统签名机制特殊性，难以验证签名厂商信誉以及其他属性

统计	编译器种类	壳种类	供应商	共享文件	隐藏文件	正版授权	未授权软件	持续活跃	偶尔活跃	持续静默	静默
按属性统计	12个	25个	36个	0个	320个	128个	0个	5000个	1.6W	4W	1.9W

统计	有已知漏洞	敏感命令创建的文件	服务执行	启动项执行	生僻技术栈	生僻编译器	生僻壳	生僻厂商	名称伪装	图标伪装	敏感端口监听
安全性统计	13个	360个	78个	60个	5个	4个	2个	4个	6个	0个	12个

国产化办公主机

操作系统：Android 11　　系统安装时间：2020/09/15　持续监测时间：5天

统计	执行体总数	可执行程序	模块	系统文件	系统应用	so文件	工具软件	办公软件	文档	脚本	CLASS文件	其他应用
综合统计	10W	1200个	7852个	3.2W	273个	6425个	62个	37个	1036个	8652个	4.1W	89个

统计	系统签名	普通签名
数字签名统计	273个	135个

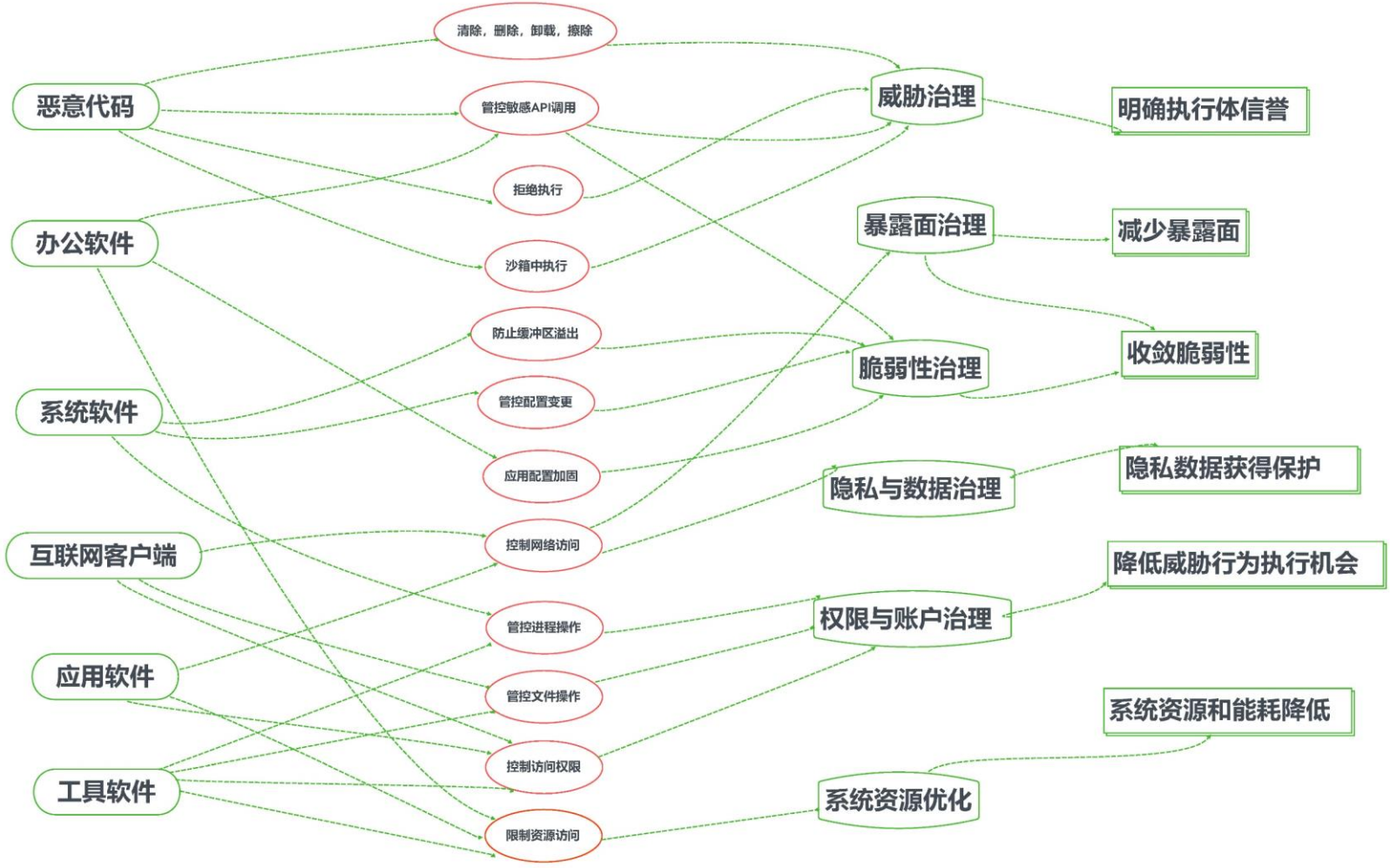
统计	技术栈	加固壳种类	大文件	普通文件	隐藏文件	图片文件	文档文件	后台运行应用	系统预置应用	第三方应用
按属性统计	4个	4个	223个	35254个	3320个	128个	365个	18个	273个	135个

统计	有已知系统漏洞	敏感权限申请使用应用	服务执行	开机启动	申请短信权限	申请摄像头权限	申请位置权限	申请sd卡读写权限	申请网络权限
安全性统计	16个	74个	65个	44个	6个	32个	18个	64个	146个

BYOD移动终端

03 执行体治理所需的能力集合与运营模式

执行体治理的关联图谱


清除，删除，卸载，擦除
管控敏感API调用
拒绝执行
沙箱中执行
防止缓冲区溢出
管控配置变更
应用配置加固
控制网络访问
管控进程操作
管控文件操作
控制访问权限
限制资源访问
恶意代码
办公软件
系统软件
互联网客户端
应用软件
工具软件
威胁治理
暴露面治理
脆弱性治理
隐私与数据治理
权限与账户治理
系统资源优化
明确执行体信誉
减少暴露面
收敛脆弱性
隐私数据获得保护
降低威胁行为执行机会
系统资源和能耗降低

执行体治理的对象任务

工作任务	工作对象	具体动作	紧急程度	工作难度	工作代价
有效处置响应已知威胁对象	恶意执行体	拒止、清除、反持久化、取证、分析、恢复、量损、溯源等	高	低	小
筛选排查可疑对象	可疑执行体	情报匹配、样本分析、流量分析、日志分析、关联拓线、异常分析等	高	中	中
塑造基础运行环境	执行环境的安全配置	可信引导链、收敛运行入口、配置基线策略、管理镜像等	中	中	中
识别与管控动态运行对象	活跃执行体	执行体基线管理、执行环境与执行行为监控、执行体微隔离策略调优等	中	中	中
识别分析判定静默对象	全量执行体	执行体元数据采集与信誉标定、执行体信誉情报批量查询、动态沙箱分析、软件组成成分分析、软件静态代码分析等	中	高	大
管理和约束供应链与软件来源	全量代码与执行体	强制签名要求、审计要求、软件物料清单管理等	低	高	大

防御能力框架下的执行体治理能力枚举

	识别	塑造	防护	检测	响应
执行体治理所需能力集合	• 系统环境识别 • 网络环境识别 • 执行体与业务支撑关系 • 用户及权限 • 系统配置 • 暴露面 • 脆弱性 • 进程行为 • 模块行为 • ……	• 系统环境策略 网络管控策略 • 主机加密环境 • 流量加密环境 • 虚拟执行环境 • 网空欺骗环境 • ……	• 系统资源访问拒止 • 服务访问拒止 • 连接创建拒止 • 连接传输拒止 • 文件创建拒止 • 系统资源创建拒止 • 写入拒止 • 执行拒止 • 加载拒止 • ……	• 系统加载点检测 • 流量环境检测 • 应用环境检测 • 数据体检测 • 执行行为检测 • 用户行为检测 • ……	• 缓解（临时下达网络管控策略、环境塑造策略） • 固证 • 主机环境处置 • 网络阻断 • 环境与数据恢复 • 持续策略调整 • ……

需要多种模块和服务支持

系统模块	*数据采集探针模块*	实现对全量资产内执行体全要素数据的采集与元数据识别，支撑上层业务的运行
	应用管控模块	基于元数据识别、信誉评估结论对执行体进行细粒度管控
	分析资源池	使用采集的执行体元数据、行为数据等内容对执行体信誉进行评估并形成相应结论
	规则管理模块（新增）	对执行体采集、识别、管控等能力的规则进行配置，使管理人员可对系统进行纠错或者调优
	情报（信誉）中心	提供海量威胁情报、信誉知识，可提升分析资源池的分析准确性
	算力和存储资源池	提供执行体治理过程中必要的算力与存储资源需求
运营服务	***人工分析服务***	协同网络安全管理人员对资产环境或者执行体信誉进行人工分析
	威胁猎杀服务	协同网络安全管理人员对发现的威胁执行体构建管控规则并执行威胁清除、固证、溯源等服务

关键支撑能力——海量恶意代码精准识别能力

有效处置检出的恶意代码，依然是执行体治理最重要的环节。

高级攻击组织未必使用0DAY漏洞和复杂的恶意代码，使用开源、商业恶意代码，劫持第三方机会的情况也十分普遍。

能够被低风险恶意代码代码感染的系统，一定可以被高级别攻击者攻陷。

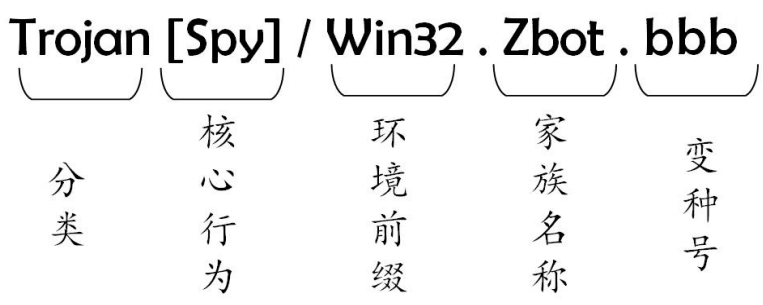
高安全场景恶意代码处置已经不是一杀了事，需要固证、分析、归因、补漏、量损、止损等动作。

文件格式	规则数
PE	41,292,256
安卓	7,918,221
脚本	1,219,258
ELF	189,332
宏病毒	141,812
溢出	6,430
其他文件	191,713

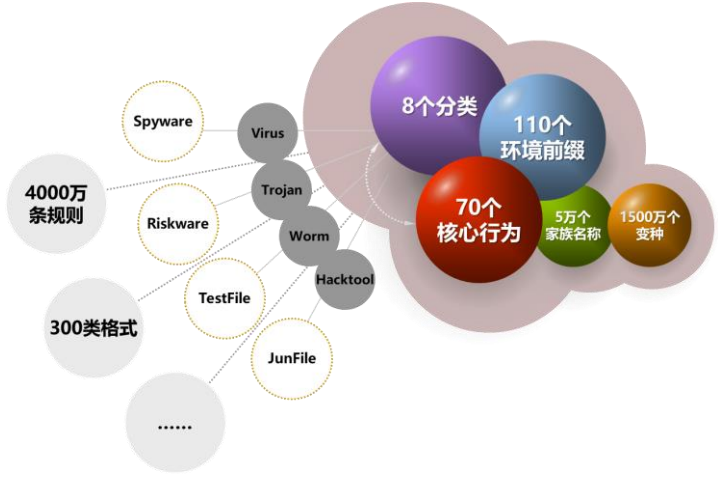
安天引擎文件规则数量

文件格式	种类数
软件数据	127
包裹	41
可执行格式	41
媒体	35
文档	31
图片	22
文本	18
脚本	9
其他文件	7

安天引擎可解析格式



安天恶意代码分类命名规范



安天的恶意代码检测能力参数
(截止2022.12.31)

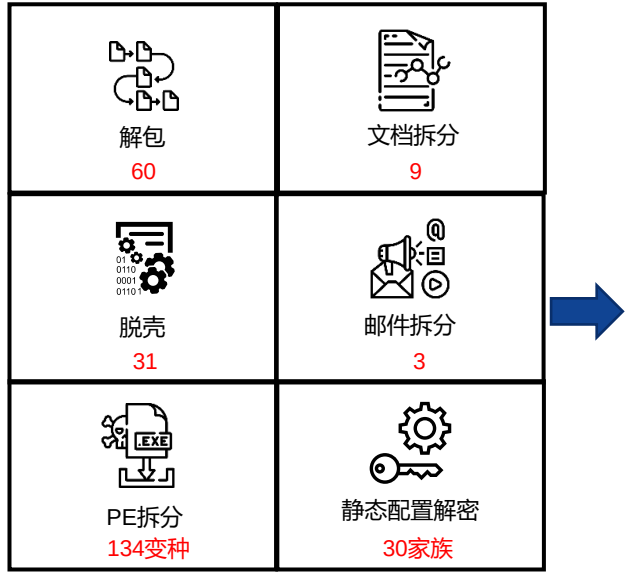
关键支撑能力——执行体的元数据提取能力

什么是元数据化

- 是指将执行体的元数据和执行体本身分开存储的过程，支持识别、分析、追踪和管理等
- 执行体的元数据包括但不限于发布者、内容、结构、成分、版本等

为什么要将执行体元数据化

- 执行体元数据化是执行体识别和精细管控的基础
- 粗糙的标识难以支撑精细化的治理

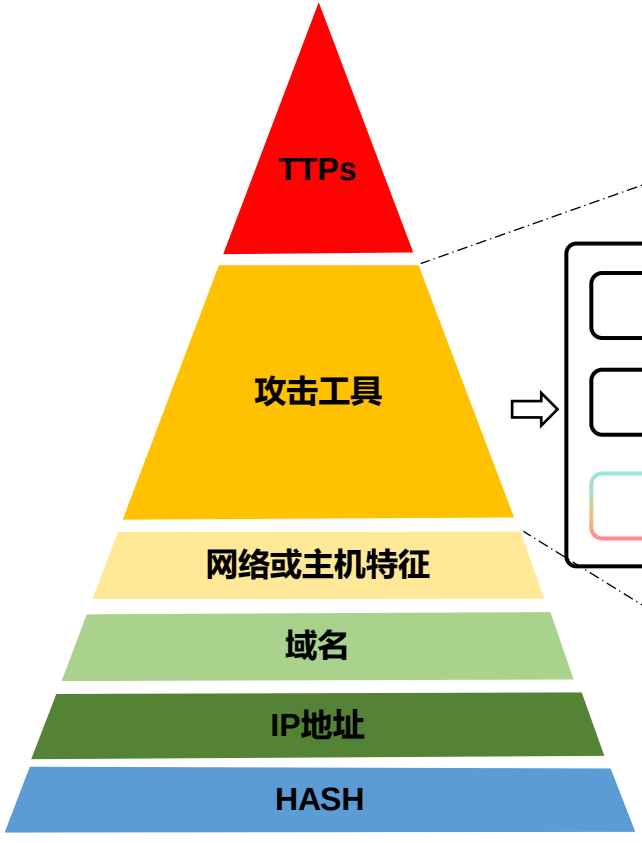


深度拆解

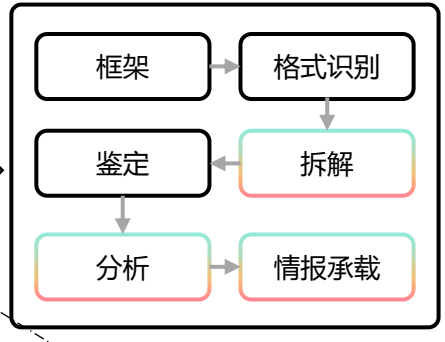
序号	类别	字段	来源	描述
1	身份信息	文档作者	属性	文档编辑者，文档的归属
		归属公司	属性	文档所归属的公司
		内嵌邮箱	正文	正文中包含的邮箱号(来源、引用、目标对象等)
2	隐私信息	语言	属性	编写文档正文所使用的语言(中文、英文、俄文等)
		包含账号、卡号	正文	正文中包含的银行卡、登录账号等隐私账号信息
3	环境信息	依赖解释器名称	属性	打卡文档需要使用的软件名称(如Office,PDF等)
		依赖解释器版本	属性	打卡文档需要使用的软件版本(如Office2016)
		依赖运行平台	属性	执行文档的平台(如windows)
4	成分信息	编码方式	结构	文档编码的方式(如Office常采用XLSB编码)
		衍生数据类型	结构	包含的所有子文件个数，以及格式类型
		内嵌的OLE对象	结构	内嵌在文档中的其他对象数据(如内嵌在PDF中的JS)
		内嵌宏代码	结构	内嵌在文档中的宏代码
		内嵌的网络资源	正文	内嵌在文档正文中的URL、Domain、IP等
5	用途信息	文档主题	属性	文档属性中标记的主题
		文档标题	属性	文档属性中标记的标题
		敏感词	正文	文档正文中包含的敏感关键词列表
6	脆弱信息	CVE漏洞	结构	包含的CVE漏洞编号列表
		包含shellcode	数据	可疑的shellcode数据
		文档破损	结构	文档结构破损，无法正常打开
7	异常信息	文档加密	数据	文档通过密码加密，需要密码才能打开
		内嵌可执行文件	结构	文档内嵌了一个可执行文件(如包含exe文件)
		内嵌钓鱼、挂马URL	数据	文档内嵌了钓鱼URL
		属性域包含代码	属性	文档属性中包含代码
		正文包含脚本	正文	文档正文包含脚本
		宏自启动	行为	文档打开或关闭能自启动宏
		内嵌脚本加密混淆	行为	内嵌的脚本进行了加密混淆
		设置安全等级	行为	内嵌的脚本修改文档安全等级
		设置保护视图	行为	内嵌的脚本修改文档保护视图
		调用Active控件	行为	内嵌的脚本调用Active控件进行系统操作
8	信誉信息	信誉度	综合	文档的信誉度，可信、恶意、未知等
		威胁度	综合	文档的威胁度，低危、中危、高危等

复合文档元数据

关键支撑能力——面向执行体的向量级情报与消费引擎



威胁情报痛苦金字塔



消费引擎

特异性向量

- 典型字符串
- 编译环境
- 签名信息
- 注册表
- 互斥量
- 通讯配置信息
- 解密密钥
- 关键代码片段等

什么是向量级威胁情报

向量级威胁情报概念由安天提出，是基于威胁检测引擎的识别和深度拆解能力承载，从执行体中抽取的的能够表征威胁行为体基因特性、具备形式化特征的深度情报。

多维基础向量

- 基础信息（字符串、编码过的二进制）
- 属性信息（格式、编译器、壳、包、版本信息）
- 结构信息（PE结构、复合文档结构、结构异常）
- 身份信息（开发者、登录ID、密码、邮箱、数字签名）
- 环境信息（注册表、路径、GUID）
- 攻击技术（执行、持久化、提权、防御规避、凭证访问、发现、横向移动、收集）

关键支撑能力——端点侧细粒度采集能力

系统环境细粒度采集必要性

是对威胁检测、固证、溯源、猎杀等上层业务的必要支撑

是进行执行体识别、信誉计算和管控的必要支撑

是实现牢固IT治理的必要支撑

系统环境细粒度采集对象

硬件/软件资产	系统环境	关键配置
文件	进程	账户
网络连接	暴露面	脆弱性
外设信息	资产性能	……

系统环境细粒度采集技术方案

采集方式	基于系统的API调用采集、利用系统WMI接口进行采集、利用系统com组件接口进行采集、获取系统内存进行采集、与系统服务交互进行采集、与系统内核交互进行采集、磁盘遍历采集
采集频率	实时采集、定时采集、周期性采集
触发机制	消息触发、驱动监控触发、变更触发、差异触发
采集策略	采集对象策略配置 对象属性策略配置 采集频率策略配置 采集触发机制策略配置 数据上报优先级策略配置

关键支撑能力——基础信誉库

基础信誉库是海量执行体的识别基础

- 客户面对场景内海量的执行体，但客户的分析人员和分析资源有限，如果没有厂商支撑识别大多数对象，用户不可能自行完成对海量执行体的识别
- 安天拥有海量执行体的信誉库，可帮助客户解决大部分网内执行体的初始信誉基线构建，全面节省客户计算的算力

关键支撑能力——本地分析鉴定和治理情报生产能力

- 能力下沉，**解决保密与安全难以兼顾的问题。**
- 情报线索拓展，发现恶意执行体后，**输出情报，补全断链。**
- 揭示执行体行为，**在受控环境中细粒度分析执行体行为**，特别是威胁行为。
- 提升攻击者对我防御体系的**绕过、预测难度。**

环境仿真

- 超11种操作系统，包含2种国产化操作系统
- 10种协议仿真

行为监控API 监控点

- 监控点总数507。
- 涵盖文件（58）、进程（52）、注册表（41）、网络(104)、服务（18）、系统（43）、反调试（3）、证书（5）、剪贴板（5）、加解密（23）、设备（3）、浏览器（9）、office（11）、内存（3）、网络管理（10）、flash（3）、其他（32）

行为分析规则及敏感行为提取能力

- 行为分析规则1115条，覆盖网络类（52）、注册表类（322）、进程类（288）、文件类（84）、其他（369）等类别。

对抗行为揭示

- 规则总数：100+
- 类别：反虚拟机、反调试、反杀毒软件等

情报输出能力

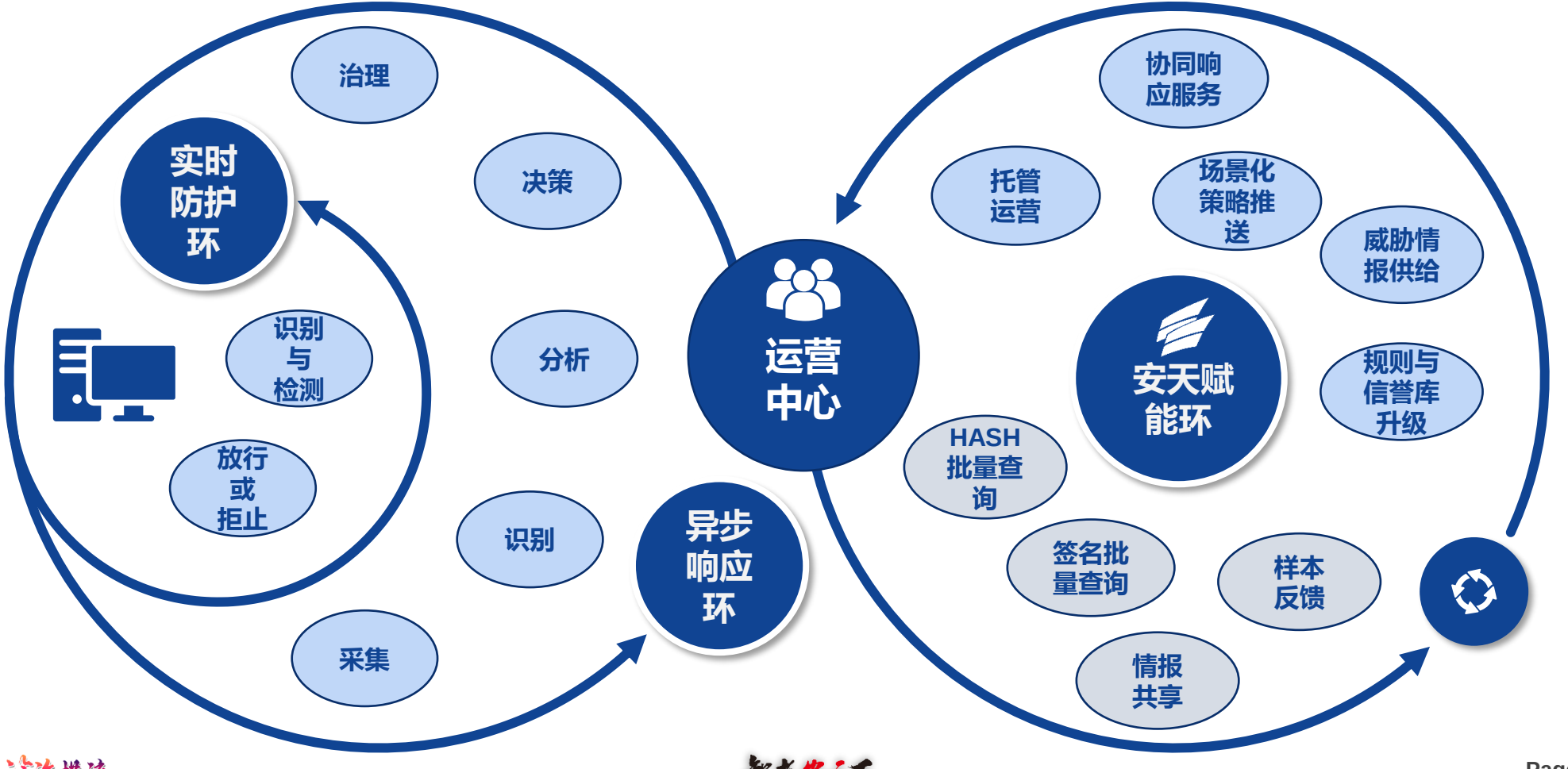
- 动态向量：大于500种
- 支持STIX等自动化交换格式输出

安天追影威胁分析系统

关键工作——构建场景化面向执行体的行动能力

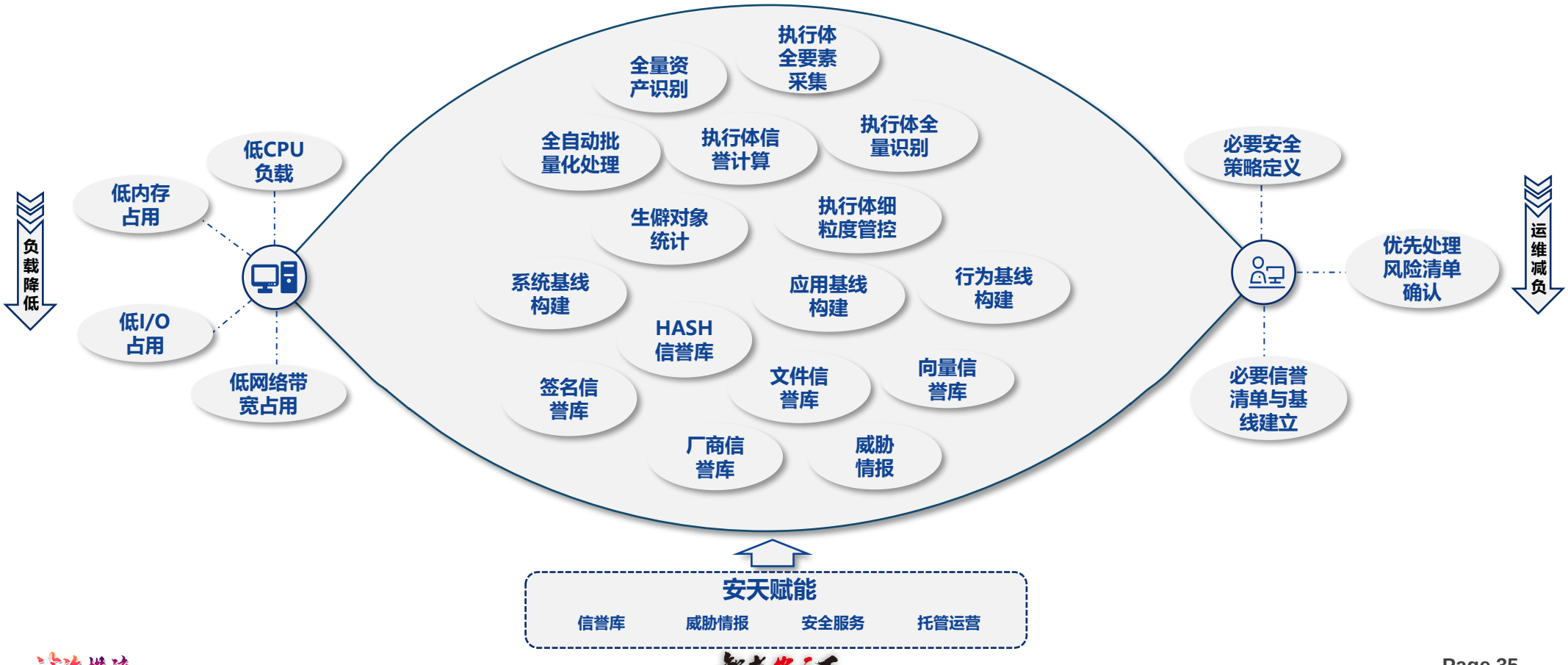
场景/安全环节	来源	元数据/计算元素	标签	判定结果（信誉）	行动
端点防护场景（实时环节）	AVLSDK引擎	/	威胁框架技战术（能力）	恶意代码检测：黑白、分类、核心行为、运行环境、家族、变种、风险等级 签名验证结果。	信任（运行管控，自动） 受限（运行管控，自动） 拒止（运行管控，自动） 清除/删除（IO处置，自动） 环境清理（IO处置，自动） 数据回滚（IO处置，自动） 固证（深度，人工） 溯源（深度，人工）
	智甲主动防御/场景采集	进程/线程信息、执行参数、执行动作、执行载体、子进程属性、父进程属性、操作账户、调用系统API、堆栈、注册表（项/键/值）、系统服务（名称/启动方式/启动对象）、任务计划、系统配置、网络五元组、用户、访问令牌、环境变量、EVENT事件、互斥量、有名管道、套接字、事件日志、代码片断、句柄、I/O占用、嵌入资源、窗口、设备、执行方式、执行时间等	威胁框架技战术（行为）、脚本执行、释放文件、创建自启动项、伪装、劫持、加密、提权、远程访问、利用系统服务、监听敏感端口、横向移动、程序注入、创建/修改账户、数据销毁、网络嗅探、网络外联等	高/中/低风险执行体	
对象分析（异步分析环节）	AVLSDK引擎（静态）	HASH、文件名、文件大小、文件格式、路径名、供应商、组件、结构、数字签名、注册表相关（项、键、值）、互斥量、GUID、PDB、编译环境、编译时间、特殊字符串、代码片段、算法、反分析、组件等	编译器类型、包裹类型、壳类型文件、结构破损、多层压缩、多层嵌套、包含敏感关键词、内嵌网络资源、内嵌文件、释放文件、自启动、键盘操作记录等	组成成分、漏洞、安全缺陷、异常类型	
	追影沙箱（动态）	运行环境、注册表行为、进程行为、网络行为、文件行为、配置修改行为、域、URL、PDB路径、互斥量、数字签名、衍生物、API调用等	威胁框架技战术（行为）、自启动、自删除、键盘记录、窃取密码、篡改服务、反调试、远控、反沙箱、漏洞利用、技战术等	高/中/低风险执行体	
联合联动与人工	智甲EDR管理中心	执行体分布统计、签名清单、厂商清单、技术栈清单、执行体活跃状态	生僻签名、生僻厂商、生僻技术栈、生僻文件、活跃执行体、静默执行体、自定义标签等	高/中/低信誉执行体	
	流量与边界侧	网络拓扑、资产通联关系	地理位置、利用VPN信道、载荷投放	正常通联、违规通联、恶意流量	

构建运营闭环


治理
实时防护环
决策
识别与检测
分析
放行或拒止
识别
采集
异步响应环
运营中心
协同响应服务
托管运营
场景化策略推送
威胁情报供给
安天赋能环
规则与信誉库升级
HASH批量查询
签名批量查询
样本反馈
情报共享

04 成熟度模型、里程碑计划

能力支撑 双端收敛


负载降低
低CPU负载
低内存占用
低I/O占用
低网络带宽占用
全量资产识别
执行体全要素采集
全自动批量化处理
执行体信誉计算
执行体全量识别
生僻对象统计
执行体细粒度管控
系统基线构建
应用基线构建
行为基线构建
HASH信誉库
签名信誉库
文件信誉库
向量信誉库
厂商信誉库
威胁情报
必要安全策略定义
优先处理风险清单确认
必要信誉清单与基线建立
运维减负
安天赋能
信誉库
威胁情报
安全服务
托管运营

不同场景环境的执行体治理工作要点比较

场景	环境和执行体特点	执行体治理相关工作	
传统政企联网终端	• 节点数量通常巨大、物理空间分散、部分便捷移动 • 应用众多、类型众多 • 用途：用于业务服务、办公服务、个人等 • 行为：本机行为、外设行为、网络行为	• 清点：尽量清点 • 基线：以系统基线为主、相对灵活 • 风险：定期治理、尽量管控 • 威胁：威胁防护为主、异常检测为辅	持续治理
孤岛节点	• 节点数量通常较少、物理空间分散 • 应用较少、类型较少 • 用途：用于业务服务、办公服务、个人等 • 行为：本机行为、外设行为 仅内网行为	• 清点：尽量清点 • 基线：以系统基线为主、相对灵活 • 风险：定期治理、尽量管控 • 威胁：威胁防护为主、异常检测为辅	定期治理
工作站节点	• 节点数量较少、物理空间集中 • 应用较少、类型较少 • 用途：用于业务服务、外设（下位机）控制 • 行为：本机行为、外设行为、网络行为 通常仅内网行为	• 清点：全量清点 • 基线：系统基线、应用基线、行为基线 • 风险：全面管控（定期升级/修复，按需缓解） • 威胁：威胁防护与异常检测/阻断并重	定期治理
传统IDC/服务器	• 节点数量较少、物理分布集中在少数的数据中心 • 应用较多、类型较多 • 用途：用于业务服务、办公服务 • 行为：本机行为和网络行为为主，较少外设行为	• 清点：全量清点 • 基线：系统基线、应用基线、行为基线 • 风险：持续的全面管控（及时升级/修复，按需缓解） • 威胁：威胁检测与异常检测并重、按需阻断/告警	持续治理
云/容器或异构场景	• 节点海量、物理分布异地异构的数据中心 • 应用众多、类型众多 • 用途：业务服务为主，办公服务为辅 • 行为：本机行为和网络行为为主，无外设行为 海量的内网（东西向）行为	• 清点：全量清点 • 基线：系统基线、应用基线、行为基线 • 风险：智能化的全面管控（及时升级/修复，按需缓解） • 威胁：威胁检测与异常检测并重、自动化、按需的阻断/告警	自动/智能化治理、持续治理

对执行体的安全治理实践（云端微隔离为例）

01

全面提升执行体可观测性

- 执行体的细粒度清点
- 执行体网络流量的可视化
- 执行体风险和威胁的检测与关联

02

持续的执行体级安全基线

- 传统系统基线外，提供执行体级的应用基线
- 采用自学习建立执行体的网络行为动态基线并持续检测

03

动态的执行体级访问控制

- 基于身份标签的执行体访问控制
- 面向业务动态迁移与弹性伸缩，自适应动态调整安全策略

04

基于执行体的精细化响应

- 基于”位置”、“环境”、“业务”、“角色”精细化圈定响应策略作用范围
- 围绕执行体执行动作的细粒度响应控制

执行体治理成熟度级别

级别	级别名称	描述	典型能力与技术举措	典型管理举措
5级	**持续优化**（主动运营持续优化）	在4级基础上，适配业务的数字化演进速度，主动调整优化指标体系，改进治理流程，持续运营规则库、基线库和模型库，持续提升治理体系的运行效率。	威胁对抗运营平台、统一威胁情报运营平台、XSOAR、XSAIM……	安全开发运维一体化指南、向量级威胁情报运营规范、威胁猎杀指南、执行体治理效能评估规范……
4级	**细致评估**（全量识别细粒度管控，指标体系量化可评估）	识别全部执行体，全面掌握执行体和执行体的行为与业务之间的支撑关系，建立信誉指标、行为指标、业务影响指标等量化指标，以指标为指引针对不同场景建立配套的管控规则库、基线库和模型库。	SAST、DAST、向量级威胁情报交换、向量级威胁情报生产……	应用静态安全分析与测试要求、应用动态安全分析与测试要求、向量级情报应用指引……
3级	**清晰可控**（分布构成清晰，执行行为可管控）	在流程运行闭环的基础上，全面掌握执行体的静态分布情况与业务应用的执行体构成，建立信誉清单，同时能够识别执行体的执行动作并依据基线进行控制。	软件供应链管理、代码成分分析、应用运行时防护、微隔离、文件分布分析、XDR、终端检测与响应、统一执行体识别引擎、SOAR、应用沙箱化运行……	软件供应链管理规范、执行体数字签名管理规范、应用上线安全检查流程……
2级	**流程闭环**（治理流程闭环）	在基本防护良好执行的基础之上，具备识别、塑造、检测、防护、响应的流程闭环，形成理清执行体、实施全面治理的基础。	安全设备管理、安全事件管理、态势感知、资产管理、统一端点安全管理、文件分析沙箱、网络检测与响应、……	执行体安全管理制度、执行体脆弱性管理流程、执行体安全运维流程……
1级	**基本防护**（执行基本防护）	以防御恶意执行体为主要目标，使用常见的安全防护技术手段，在安全能力得到良好维护的情况下，对常见的威胁可以有效防御。	恶意代码查杀、桌面管理、网络入侵检测/防护、Web应用防火墙、网络防火墙、主机防火墙、网络访问控制、云工作负载防护、……	企业风险评估、安全监测、预警通报、应急演练等常见安全管理规范、制度和流程……
0级	**无防护**	以各系统自带安全机制进行防护，无明确的执行体治理目标和防护举措。	/	/

渐进建立执行体治理的里程碑计划

安全管理

- 执行体治理管理规范、指标体系与操作规程开发

技术手段

- 终端检测与响应
- 统一工作负载防护
- 执行体微隔离管控
- 动态沙箱分析
- 企业软件商店等
- 网络检测与响应
- 网络准入控制等
- 扩展检测与响应
- 安全配置管理
- 执行体治理效能评估
- 应急处置与恢复
- 知识与情报运营
- …

里程碑计划

- 里程碑4：建立执行体与执行环境基线
- 里程碑5：形成全量识别与细粒度管控能力
- 里程碑6：持续评估与优化
- 里程碑2：构建执行体采集闭环
- 里程碑3：完善执行体治理流程
- 里程碑1：评估执行体治理成熟度

图例

- 系统侧
- 网络侧
- 运营侧

总结

- 执行体一直是网络安全防御和治理工作的最重要对象之一。
- 执行体治理的**成熟度**，是安全建设水平的**重要度量指标。**
- **执行体的全量识别和精细管控**是高价值防御场景未来达成的安全运营状态。

网络空间威胁对抗防御技术研讨会
暨 第十届安天网络安全冬训营

沧海横流

沧海横流 方显英雄本色

安天冬训营 wtc.antiy.cn