ANTIY 安天实验室

AVER反思三部曲之三

向左走、向右走

——AV方法论的思考和重建

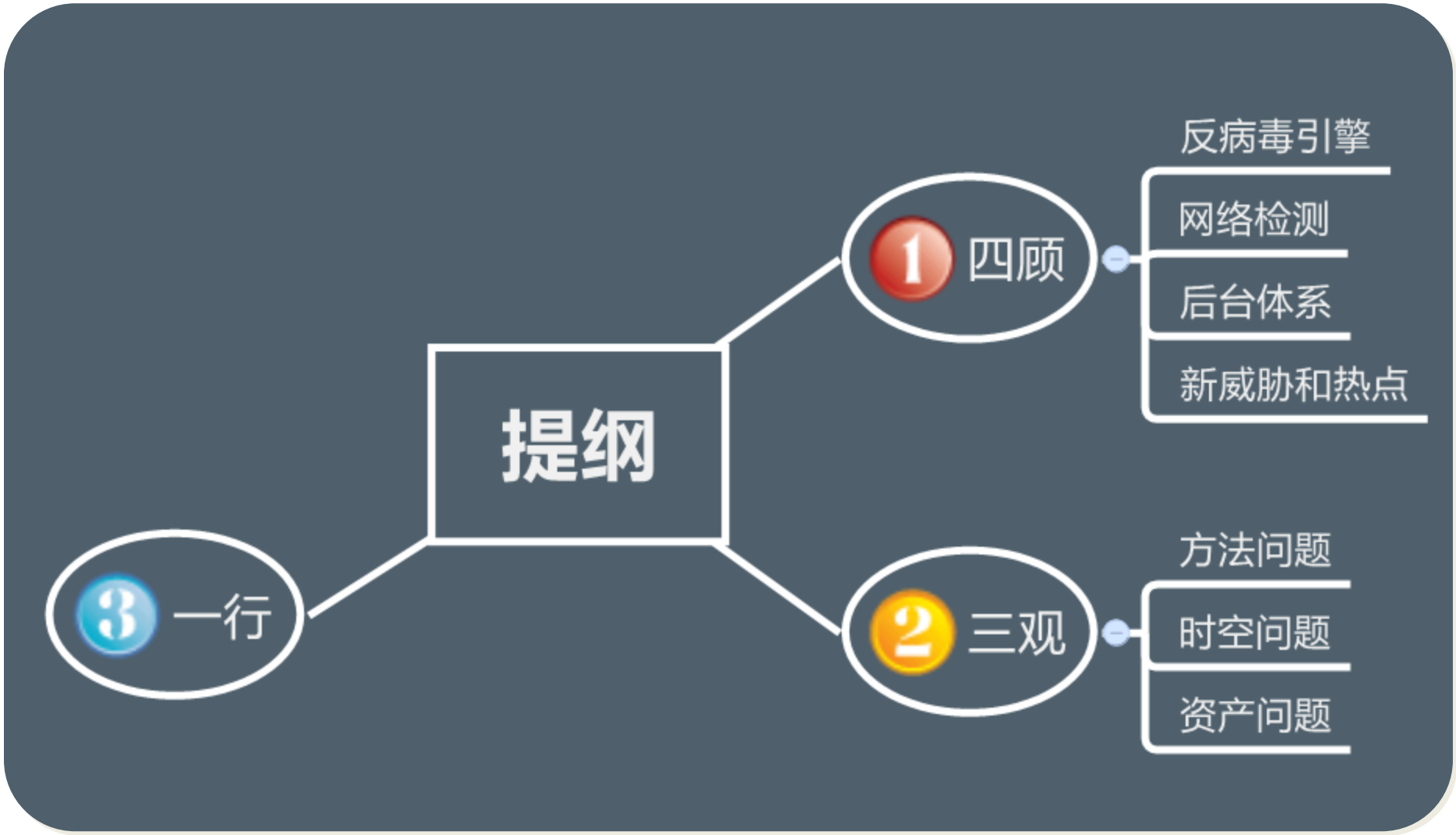
安天实验室 肖新光（江海客）

2013.ISF.上海

个人简介

- 中国公民：肖新光
- 安天实验室成员：seak
- 互联网网民：江海客
- 反病毒老兵（Since 1994）

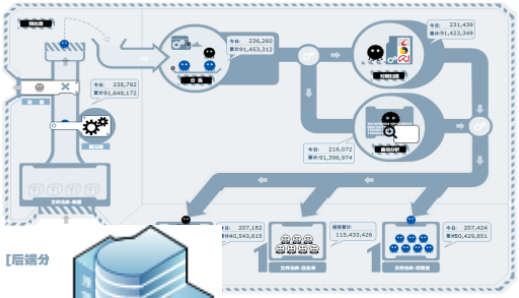
提纲



四顾：回顾AV技术的四个重要发展阶段

传统的AV的场景

后台体系



AV厂商

文件与事件采集

规则与模块分发

用户

网络测检测

反病毒引擎

主机

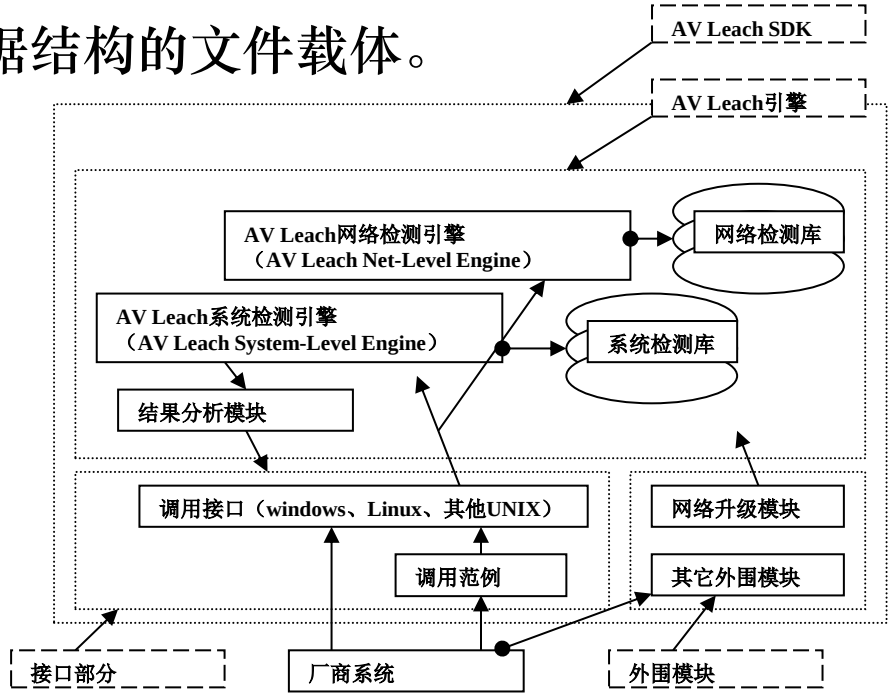
反病毒引擎的成熟（1987~1996）

⊙ 反病毒引擎

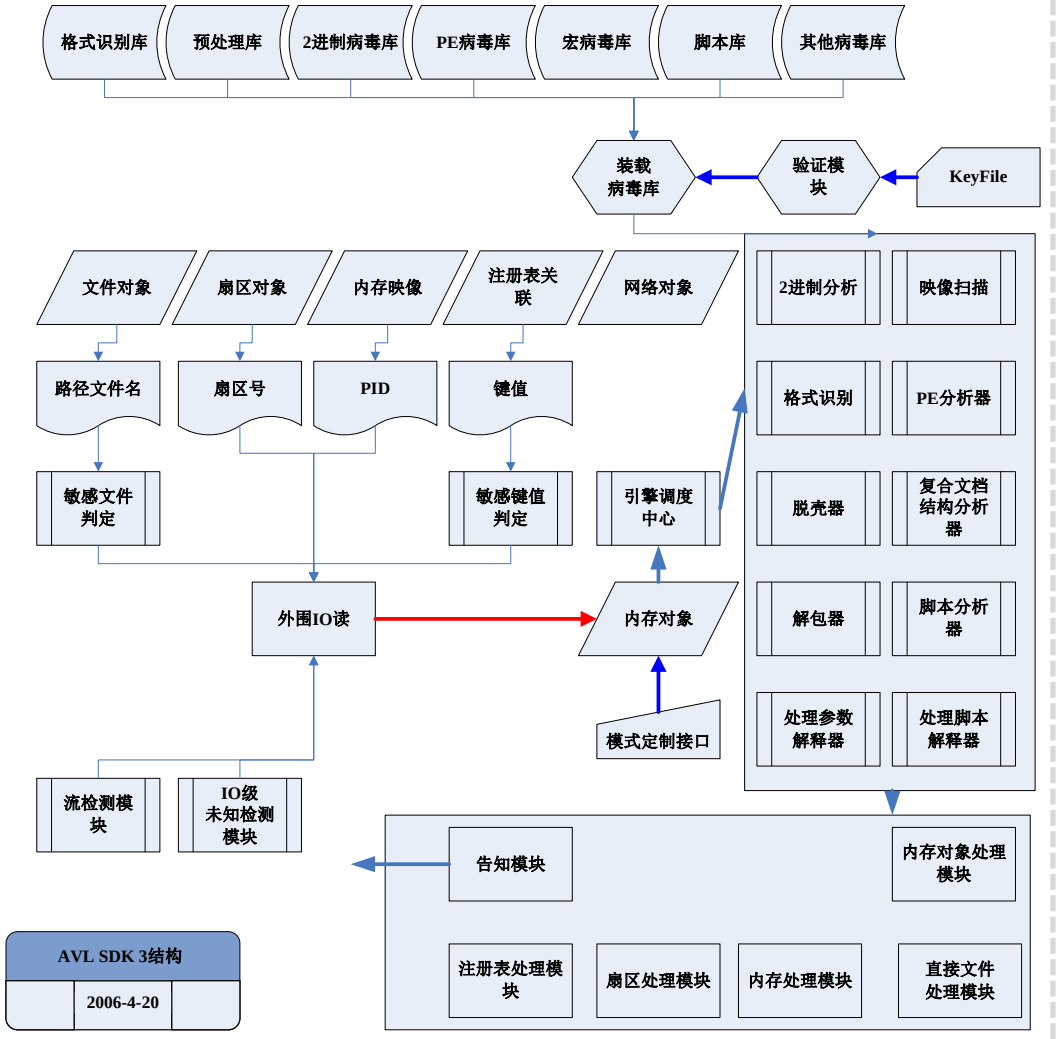
– 依赖于对应的可维护的数据结构，对待检测对象进行病毒检测和处理的一组程序模块的统称。

⊙ 病毒库

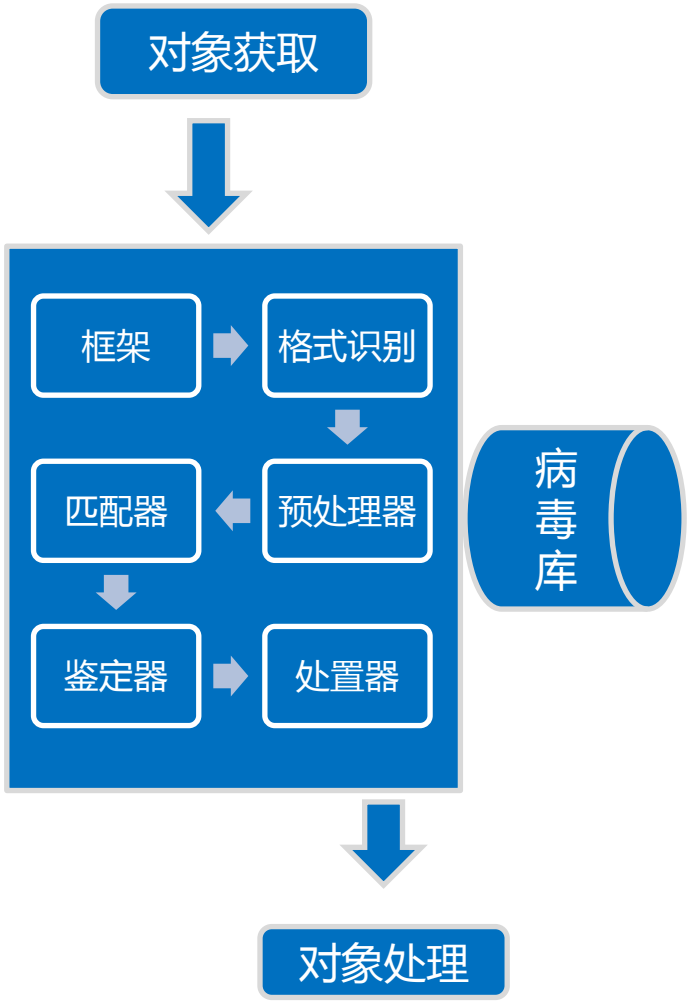
– 反病毒引擎所依赖的可维护数据结构的文件载体。



引擎的内部模型抽象化

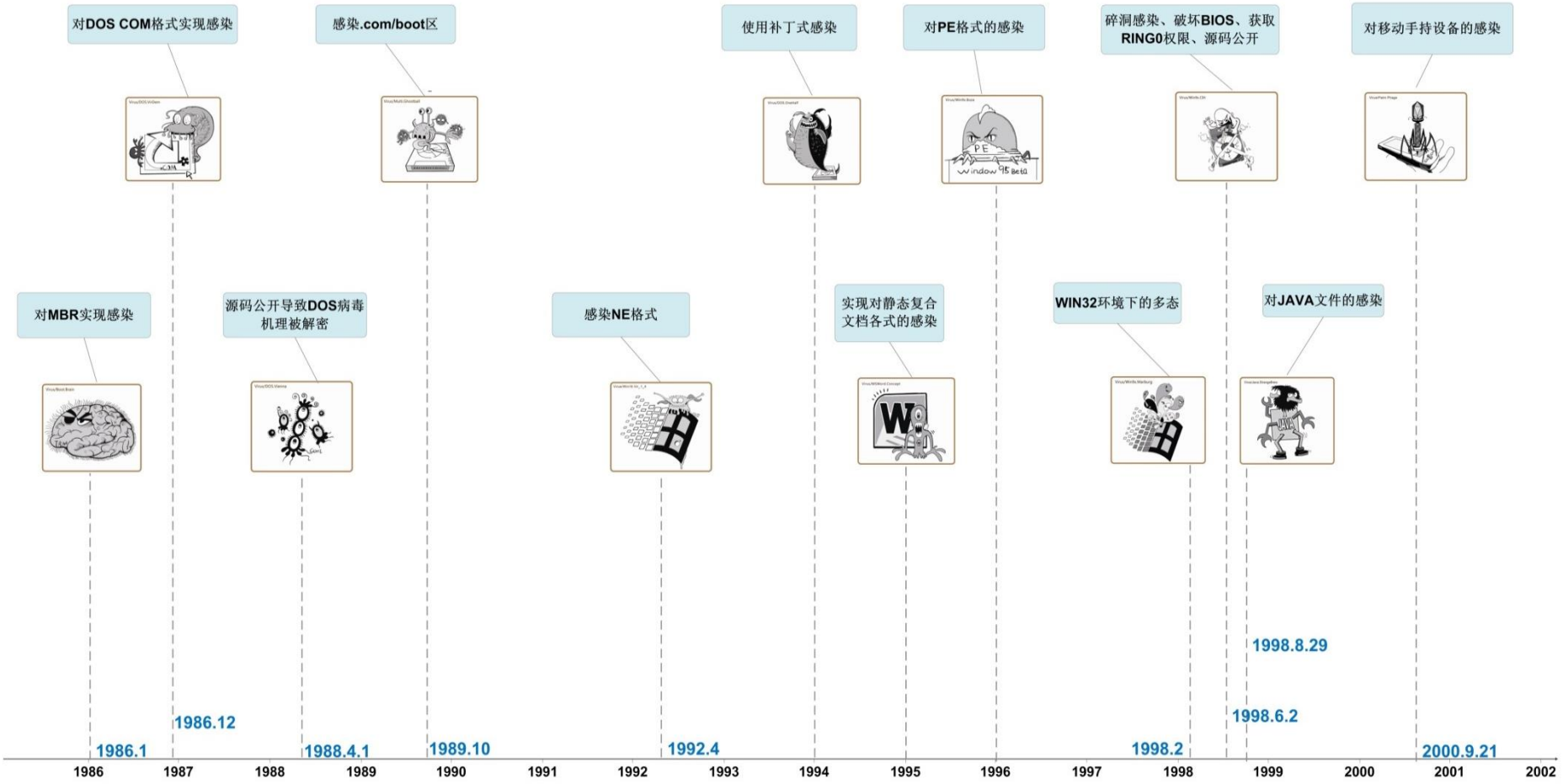


反病毒引擎的模块结构

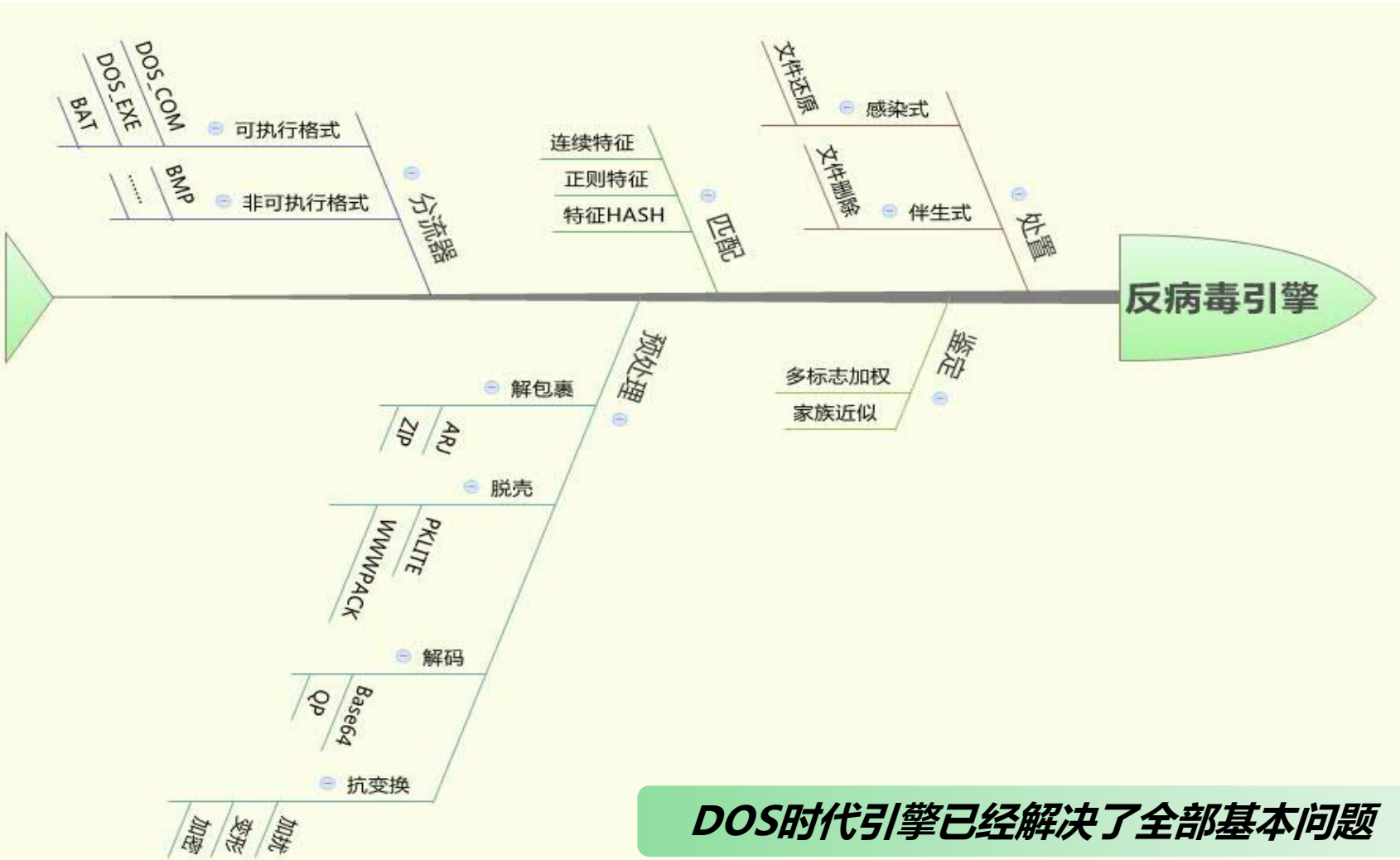


反病毒引擎的抽象结构

反病毒引擎主要靠感染式驱动成型



引擎的基本结构已经延续近20年

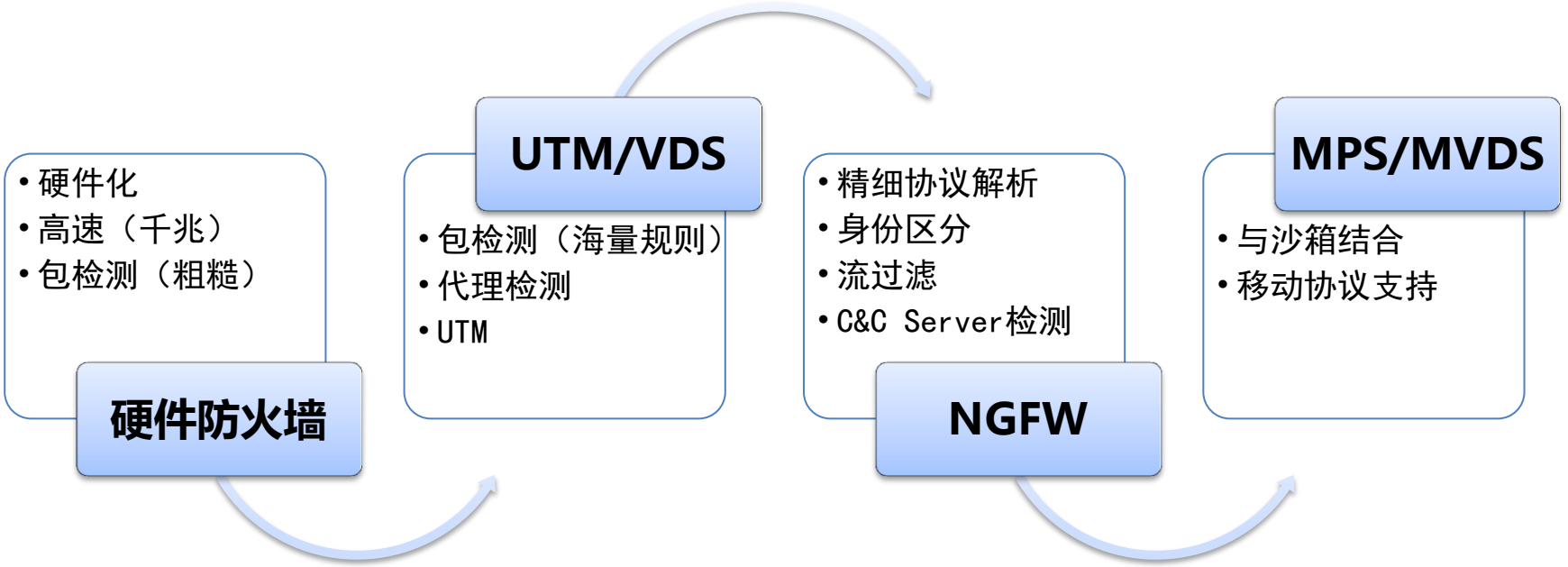


DOS时代引擎已经解决了全部基本问题

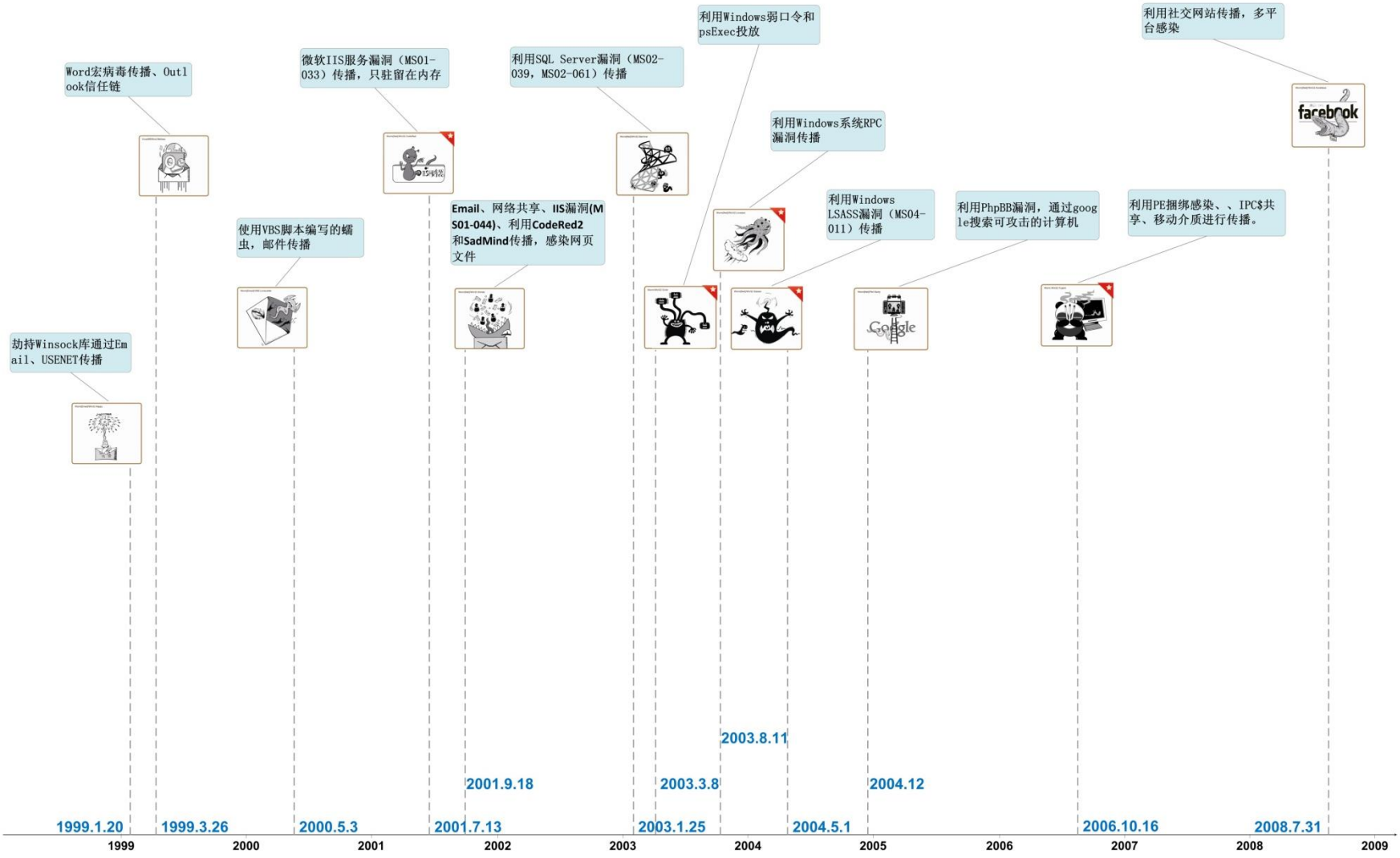
反病毒引擎的基本总结

- 反病毒引擎始于**80年代末期**对小程序的整合，成熟于**1996年**，以能够有效应对复杂结构的可执行体PE和复合文档Office为代表。
- 以**归一化**为基本方法，以**黑名单识别**为基本原理，辅助了少量白名单使用（反误报）
- 以**文件数据**为基本对象（把其他非文件数据转化为文件）
- 以**格式识别**为先导和**建立归一化分支**的前提，并为每个分支配置一组或者多组可对齐的**规则数据**。

网络检测技术的成熟（2001～2005）



蠕虫是驱动网络检测机技术发展的主要动力



网络检测技术总结

（恶意代码）网络检测技术始于2000年**初期**，至今仍在演进和发展，但其基本方法在2005年就已经成熟，其技术的发展演进基本是靠蠕虫的发展推动的。

从**基本检测对象**来看，可以分成对载荷（病毒体）的检测、行为的检测和源的检测（C&C、挂放马地址）。

从其**走向**来看，一方面不断细化检测粒度，增加检测维度，另一方面则不断应对更高带宽。

后台处理技术的成熟（1995～2010）

手工作业阶段 → C/S管理阶段与规则的自动化提取阶段 → 人工任务流水线 → 海量自动化分析流水线

1988 | 1995

- **产业背景**：操作系统和软件规模本身在初级阶段
- **核心挑战**：最基本的奠基问题。
- **主要成就**：反病毒最基础的形式化

1995 | 2002

- **产业背景**：局网应用成熟、Internet发展
- **核心挑战**：规模和复杂度的增加
- **主要成就**：现代反病毒引擎的成型、特征自动化提取技术（针对非感染式恶意代码）的成熟

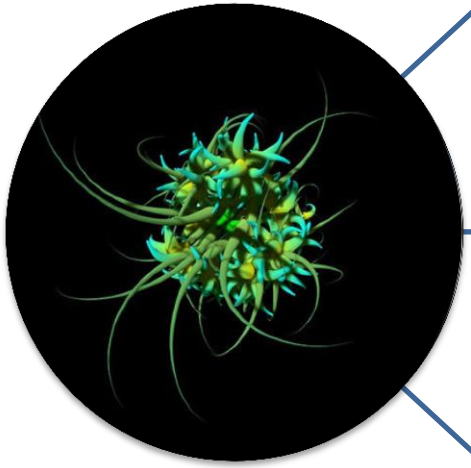
2002 | 2005

- **产业背景**：操作系统日趋复杂、网络主渠道、应用大发展
- **核心挑战**：辨识压力超越处置压力
- **主要成就**：解决分析员作业和样本管理问题

2005 | 2010

- **产业背景**：网络经济大发展催生地下经济体系，网络计算、云计算、虚拟化成熟
- **核心挑战**：样本和正常应用都以几何级数增长
- **主要成就**：解决海量样本的自动化判定问题

现代流水线的目标（ WAR ）



W

- When
- Where
- Who

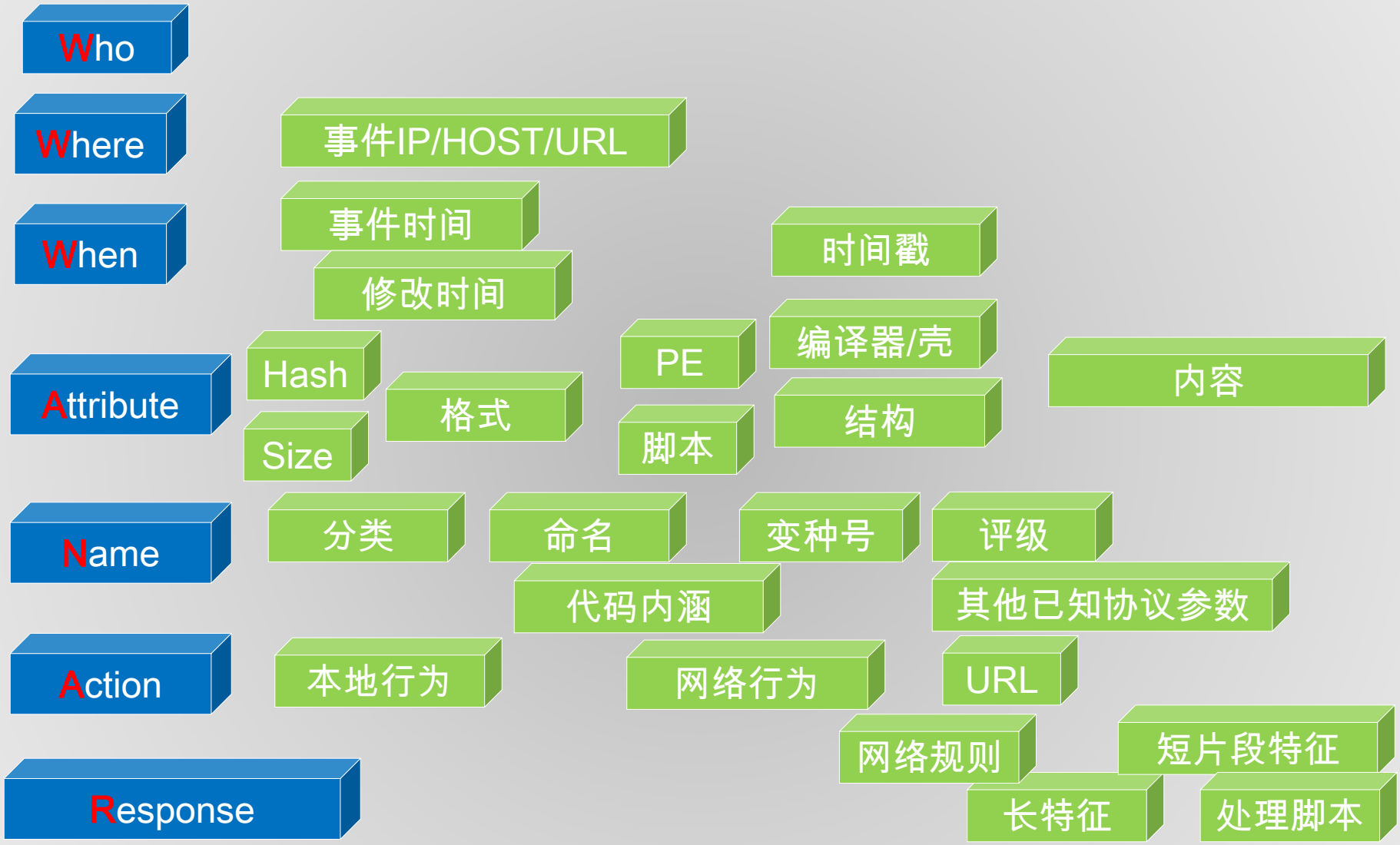
A

- Attribute
- Action
- Name

R

- Response

属性提取过程


Who
Where
事件IP/HOST/URL
When
事件时间
修改时间
时间戳
Attribute
Hash
格式
Size
PE
脚本
编译器/壳
结构
内容
Name
分类
命名
变种号
评级
代码内涵
其他已知协议参数
Action
本地行为
网络行为
URL
网络规则
短片段特征
长特征
处理脚本
Response

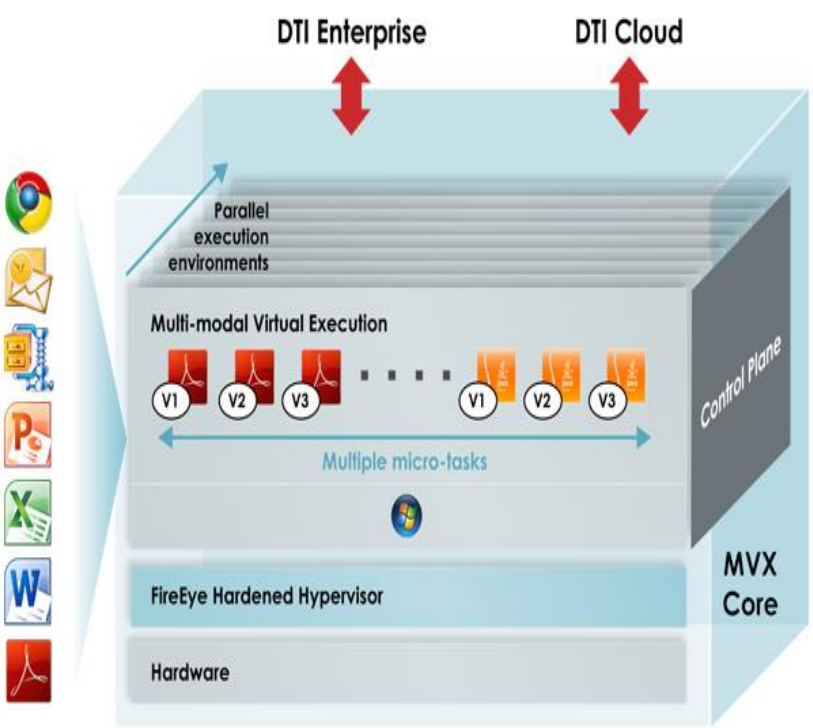
后台处理技术总结

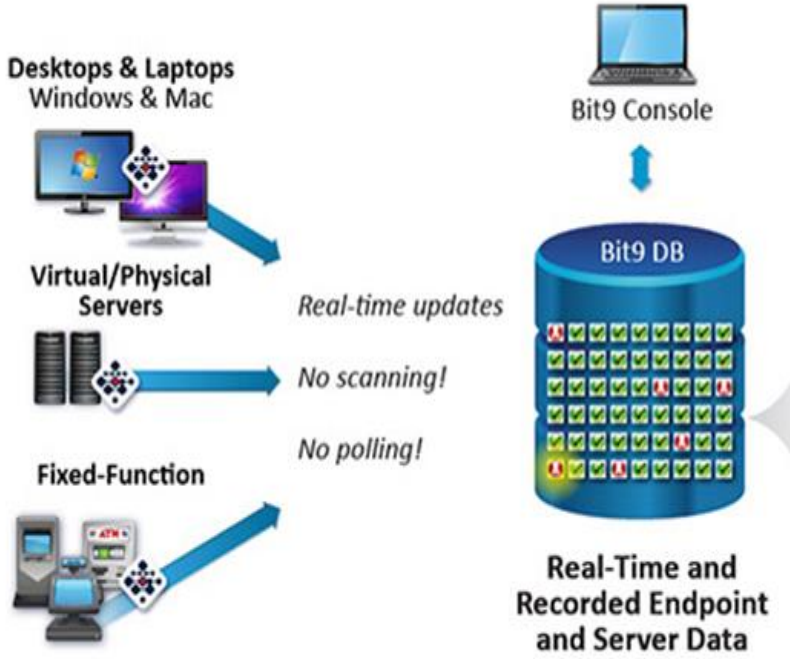
传统恶意代码后台处理是基于**多环节提取**判定思想，基于静态可分析性、行为可复现性、和厂商判定结果间的**交叉信任**。

其**根本思想**是将一个庞大的分布式计算资源的小时间片分配给海量待分析数据中的每一个样本。

后台分析体系经历了几代的发展，但对其发展的最大推动力是木马为代表的恶意代码数量的大膨胀。其基本模型在**2006年前后**确立，当前版本思想在**2010年前后**成熟。

对APT和新威胁的应对（2010～今）



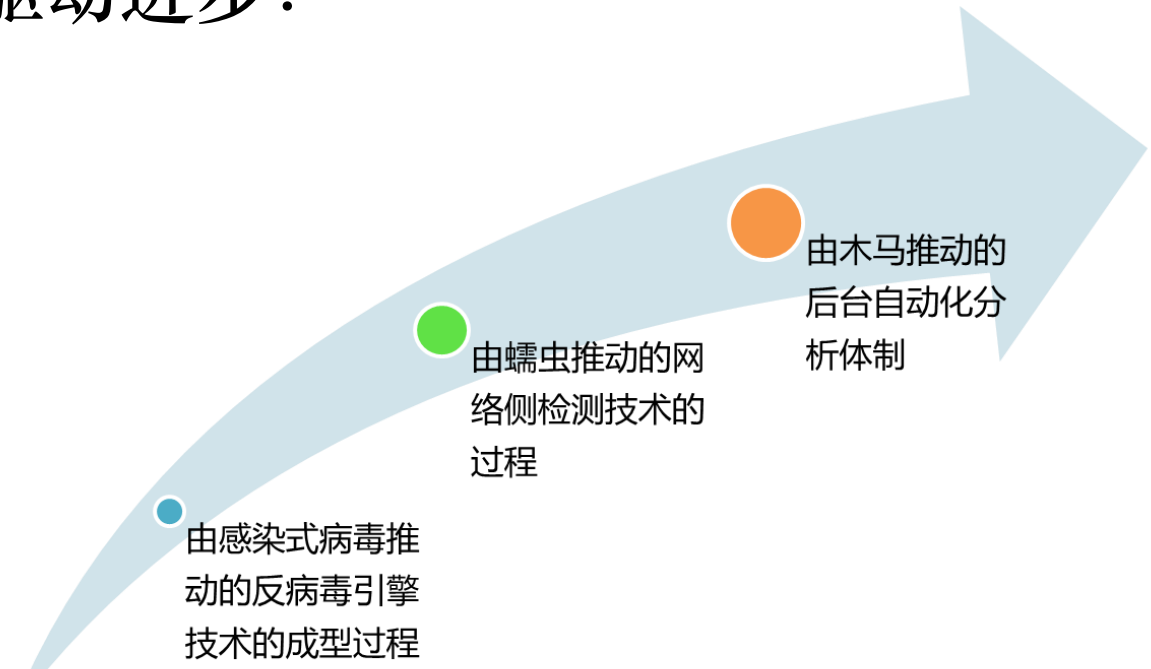


传统检测设备与沙箱结合
（Fireeye倡导）

白名单+安全基线
（Bit9为代表）

总结

⊙威胁驱动进步！



⊙这些设施和经验都是我们应对未来威胁的基础，但由于威胁方式的变化，我们需要警惕那些**已经过时**的**焦虑**、**经验**和**方法**。

这是反病毒的三观，不是世界观、人生观、价值观，而是对反病毒的方法、时空和资产的思考。

三观：关于反病毒基本方法的再思考

为什么需要反思三观

WHY？

传统AV的逐渐壮大是建立在成熟体系和庞大资源的支撑之下，其在前置经验、团队规模、计算能力方面都优于单一的攻方（VXER），是过去模型的前提假定。

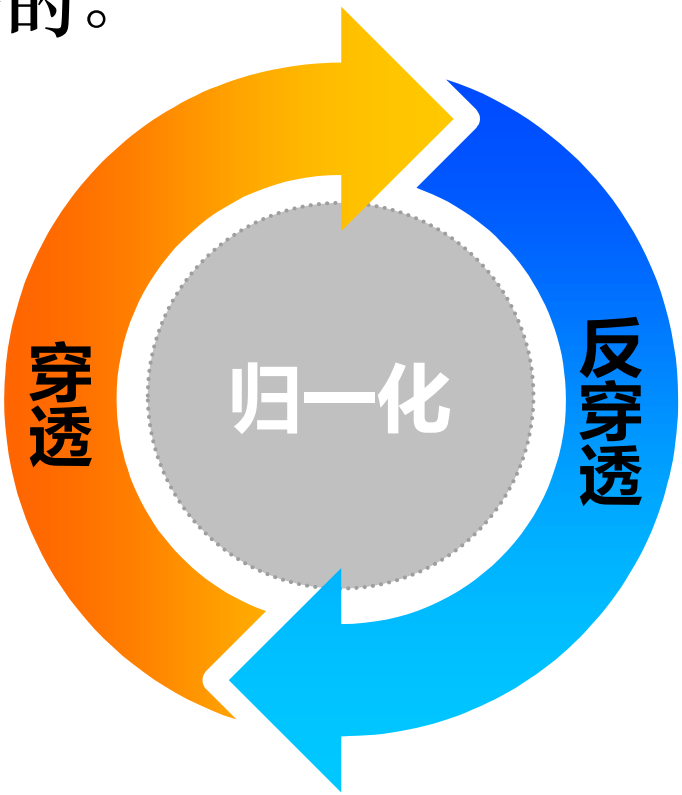
其以检测查杀攻方散布物为基本目的，通过技术手段提升了攻方的成本，并针对所捕获的样本，比攻方付出更小的资源成本和更短的时间代价，完成分析判定。

但这些前提假定在APT时代均发生了改变。

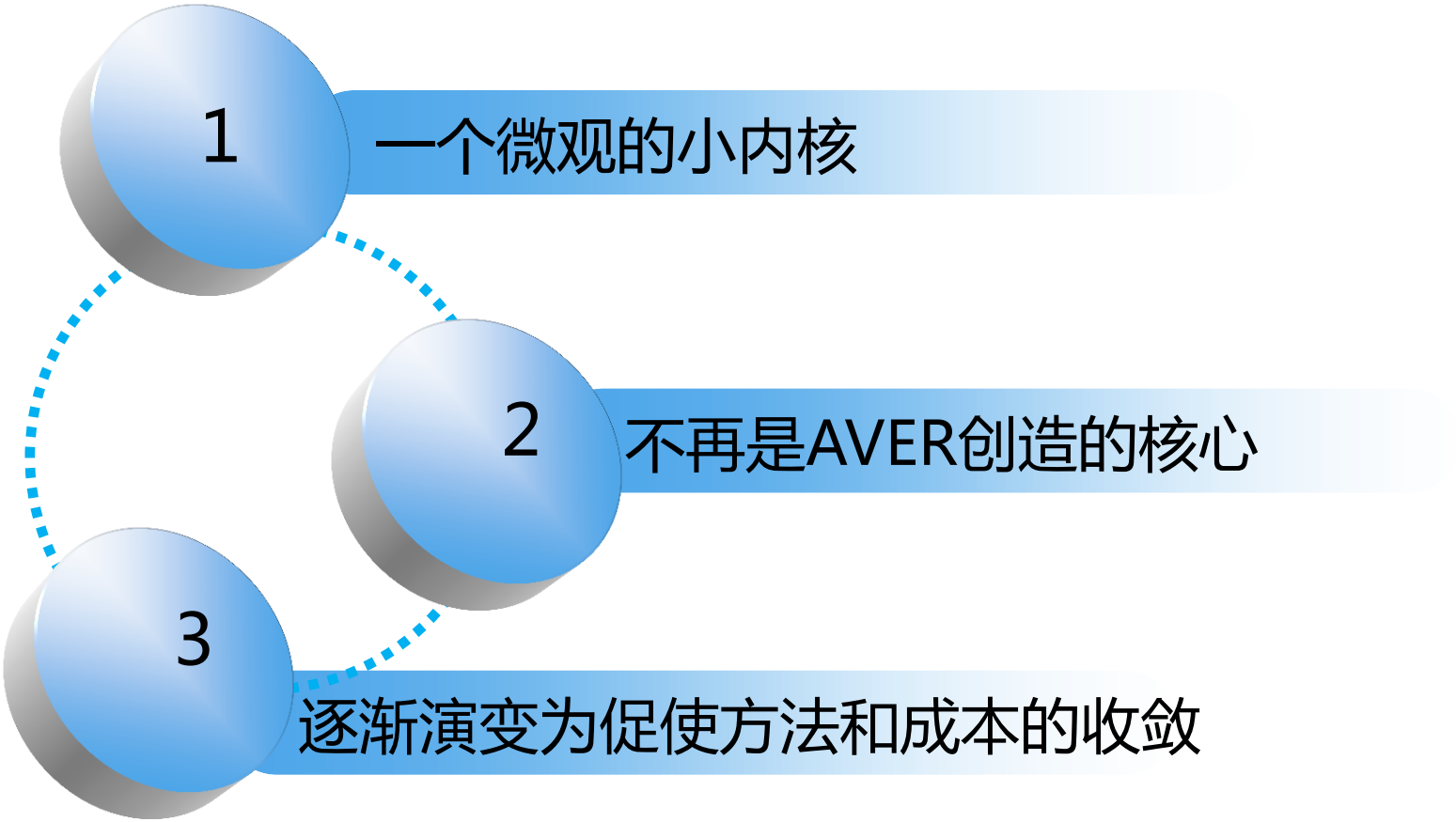
方法思考：归一化

⊙ 归一化是传统反病毒技术的精华和最重要的思想，而病毒反病毒对抗的核心也是围绕着归一化的穿透与反穿透展开的。

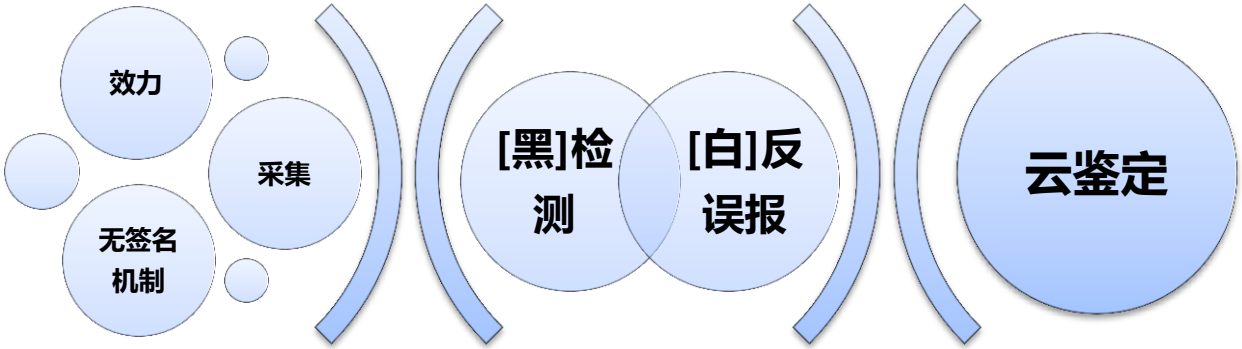
- 对规则的穿透
- 对分支的穿透
- 对链条的穿透



方法思考：关于归一化的对抗



方法思考：黑名单与白名单



以黑名单为起点是反病毒技术的唯一可行选择

从误用检测、到信誉鉴定到安全基线，AV的整体辨识重心正在发生变化

方法发思考：网络的异常检测与误用检测

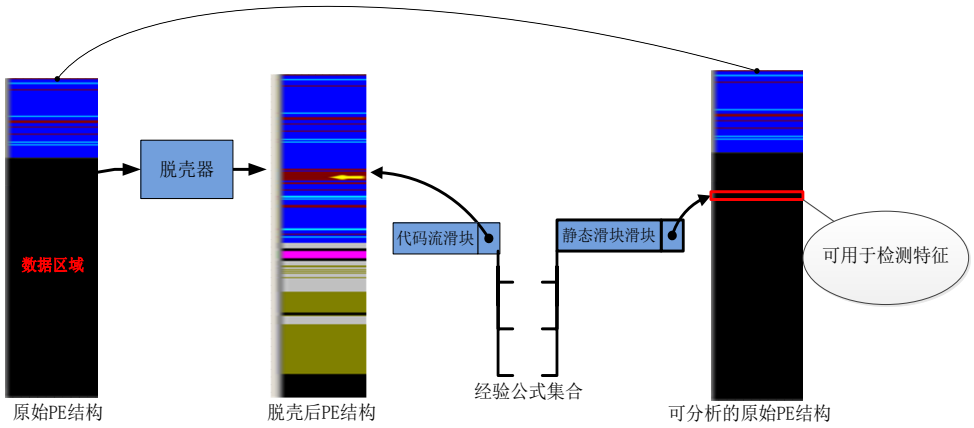
异常检测

基于其具有流量和拓扑层面的显著影响

基于在载荷分析并提取后，载荷依然被重放

误用检测

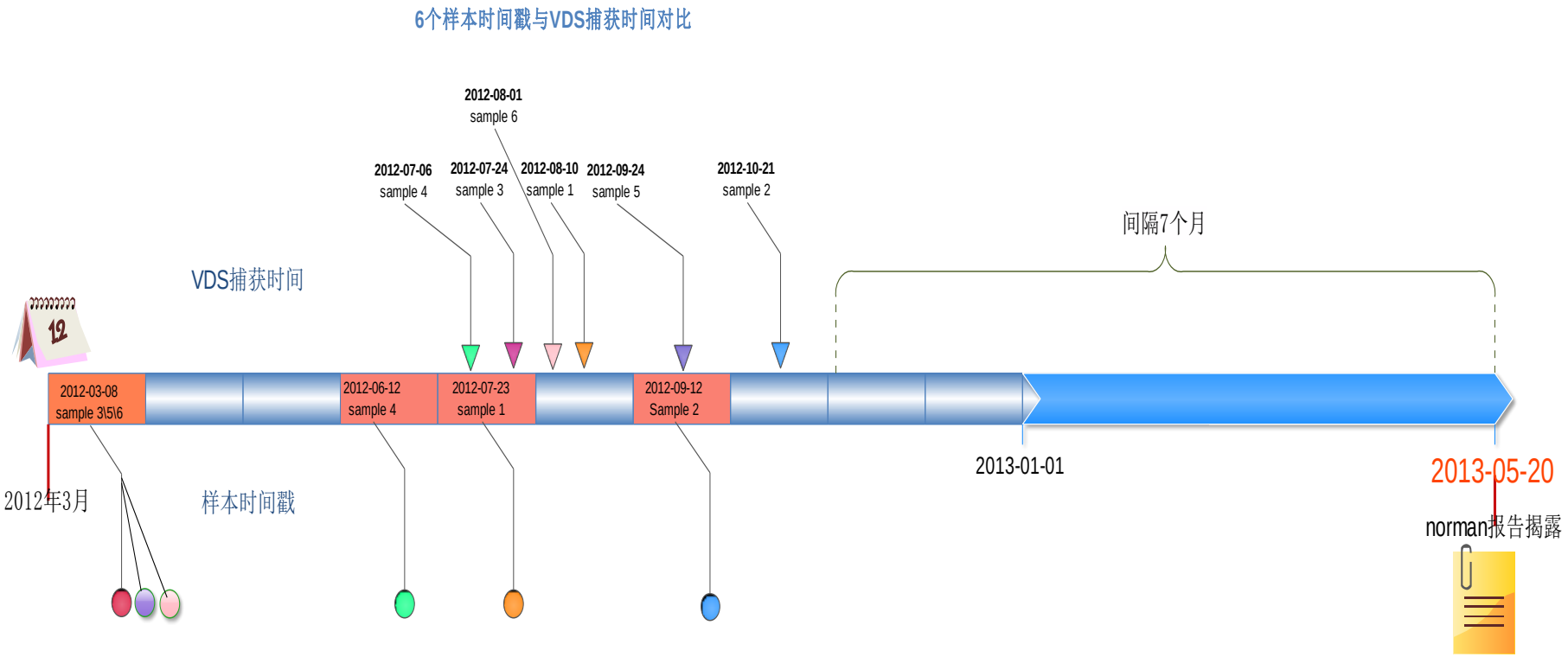
时空思考：传统的后向性



特征提取

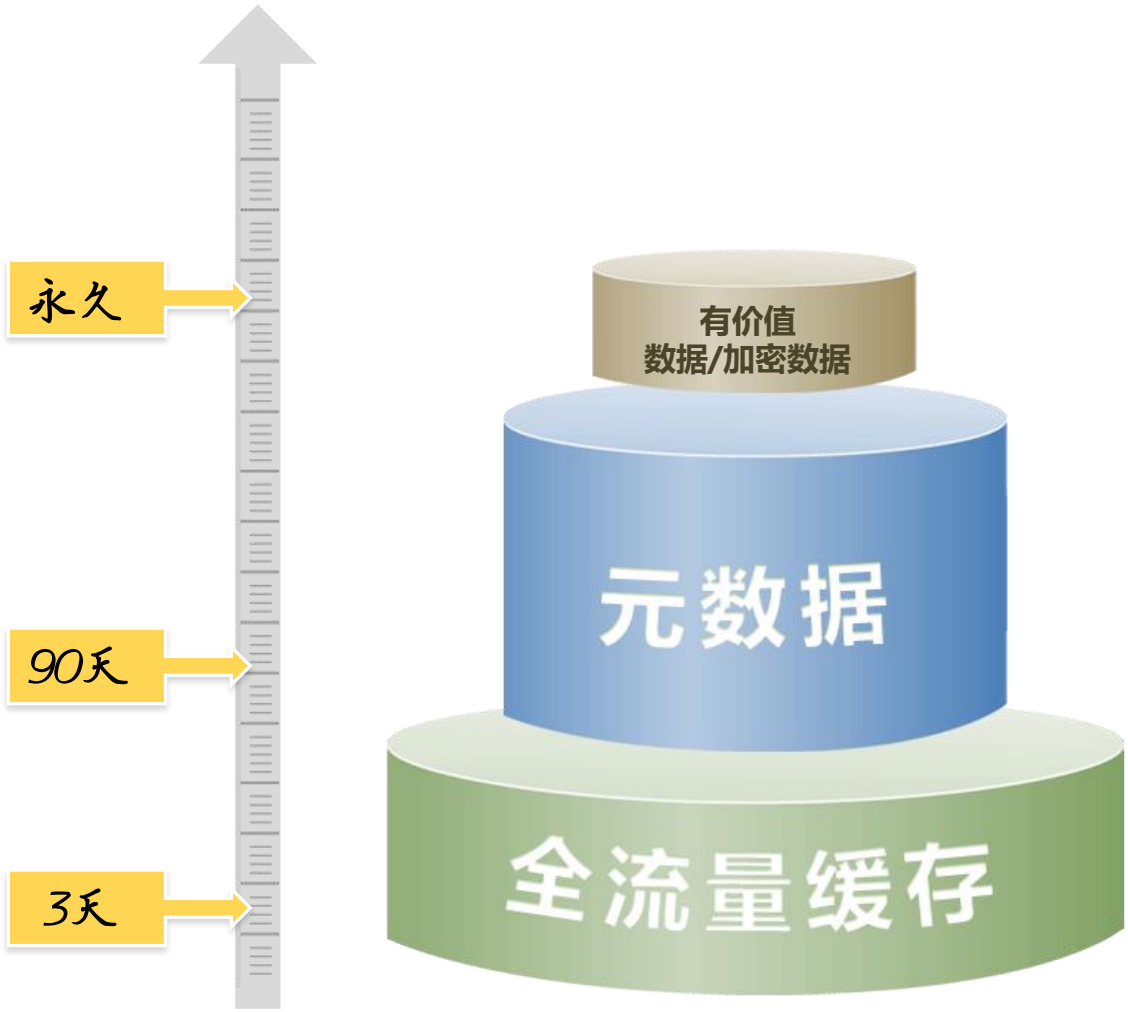
产品生效

时空思考：向前、还是向后



安天捕获Hangover时间前6个样本的情况

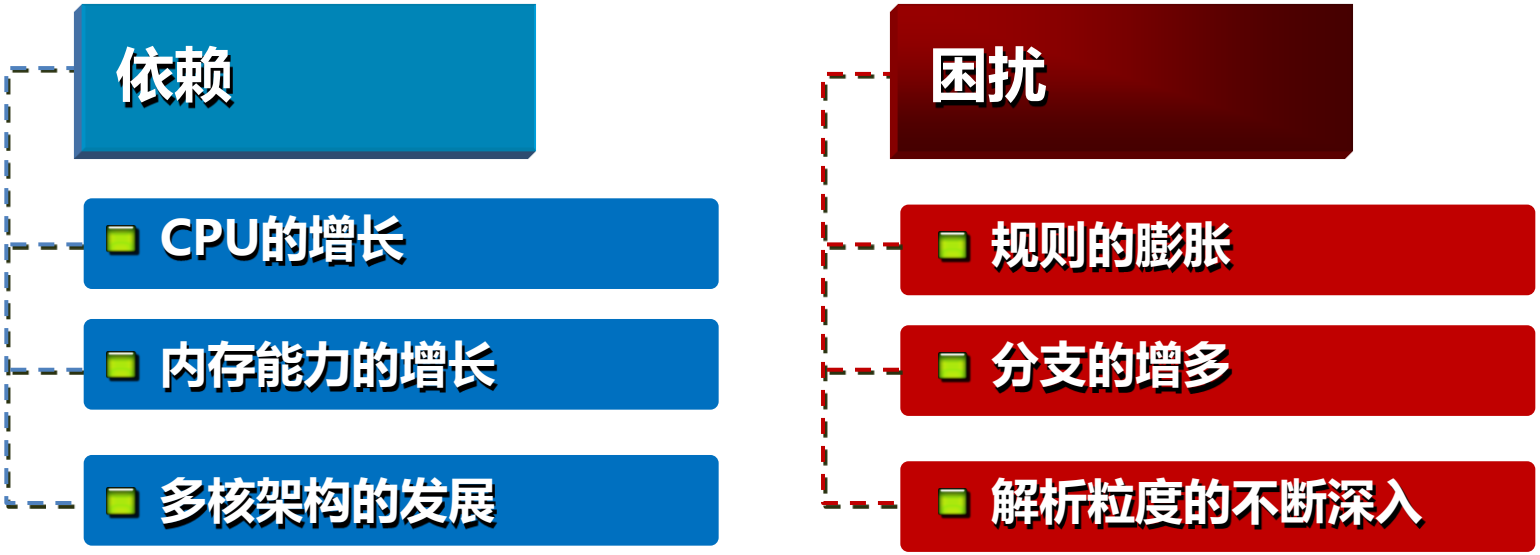
时空思考：向前需要存储能力的支持



从GCHQ的Tempora计划缓存策略看存储策略的价值

时空思考：实时还是异步

⊙传统引擎长期面对检测效率和检测效力的纠结



⊙与沙箱结合，其本质可以定性是一个异步过程，但也通过回馈给实时环节C&C Server List和URL规则的方式，改善了响应能力。

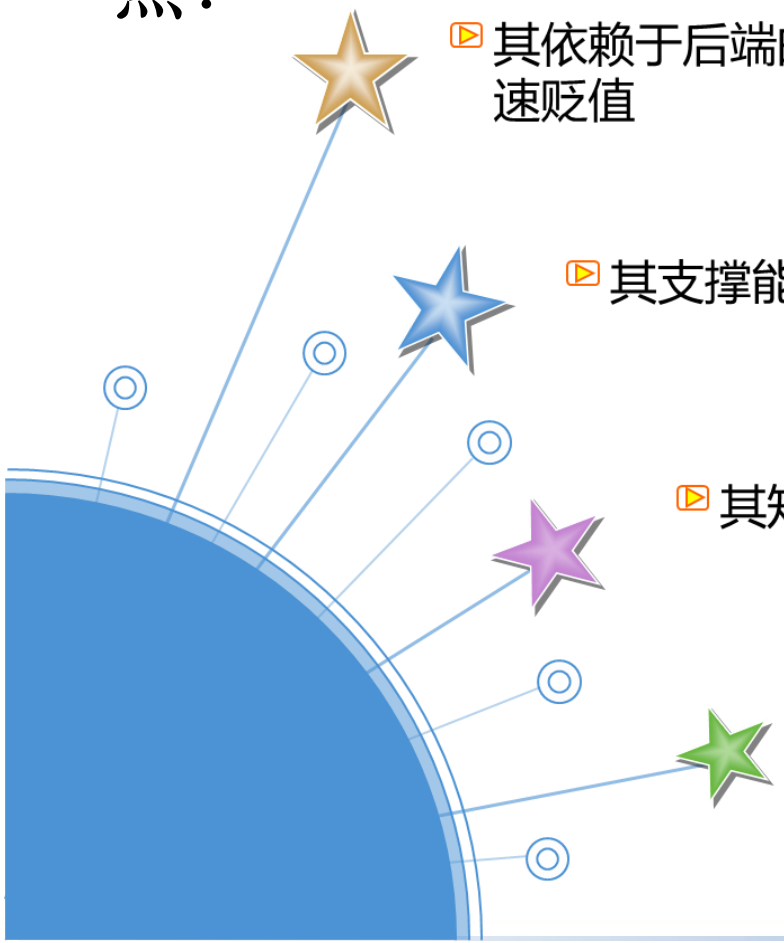
时空思考：关于部署位置

“部署位置其实就是根据个体的统计，尽量均匀，然后考虑实际情况”
——哈工大-安天联合响应小组的研究

- 在有限个节点部署的情况下实现感知的前提：
  - 蠕虫是无限制扩散的
  - 蠕虫本身是可以感知的
- 我们是否回到地缘性时代?

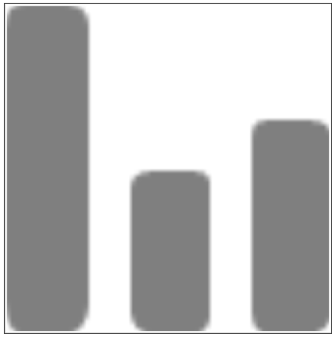
资产思考：公有与私有

⊙传统AV产品是一种资产，但其是同时具有下列特点：



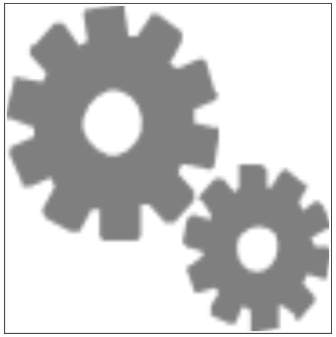
- 其依赖于后端的支撑能力，脱离后端后，其迅速贬值
- 其支撑能力即是安全保障，也带来泄密风险
- 其知识是完全为厂商所有的，且是黑箱化的
- 其自身不会增加用户的计算、存储、传输能力，相反可能会消耗用户的相应能力

资产思考：公有与私有（2）



沙箱与流量设备结合、安全基线，其反应了手段前置化、知识个性化的趋势。

其本身将原有厂商能力转化为用户个体能力，将厂商黑箱转化为用户白箱。



从资产观上看，这是用户安全资产模式的变化。其将原有厂商独占的安全资源和手段，变成用户安全资产的一部分，同时降低了在不依赖厂商支持的情况下的资产贬值速度。

本章小结

- 在APT时代，原有的AV基本方法、时空观、和资产观都在发生变化。
- 这种变化将深远的影响到整个安全的未来，以及应用的未来……

我们是孤独的反病毒引擎研发者，一人前行、一行人同行、一意孤行……

一行：关于应该做什么，关于我们在做什么？

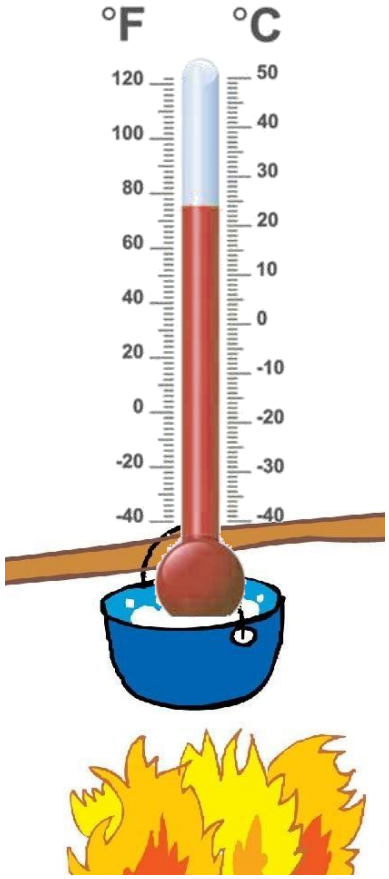
继承？还是扬弃？

- 一场有趣的，对白名单的争论。
- 虚无主义者，拒绝进行任何有效的实践，一旦其幼稚的猜测有命中的趋势，就跳出来以“语言家”自居。
- 对于费尔巴哈批判黑格尔的方式，马克思评价说：“费尔巴哈象一个糊涂的老太婆，在给婴孩洗了澡后，把婴孩和脏水一块泼到门外去了”

熟为脏水？熟为婴孩？

- 黑名单的独有价值依然是不能白名单取代的。
  - 高速
  - 精准
  - 双向孤立
- 沙箱并非万能
  - 对PE的对抗沙箱反而没有优势
  - 沙箱的先天优势是对抗格式溢出
- 白名单不只是数据

我们过去做的（AVL SDK 2.0）



高速

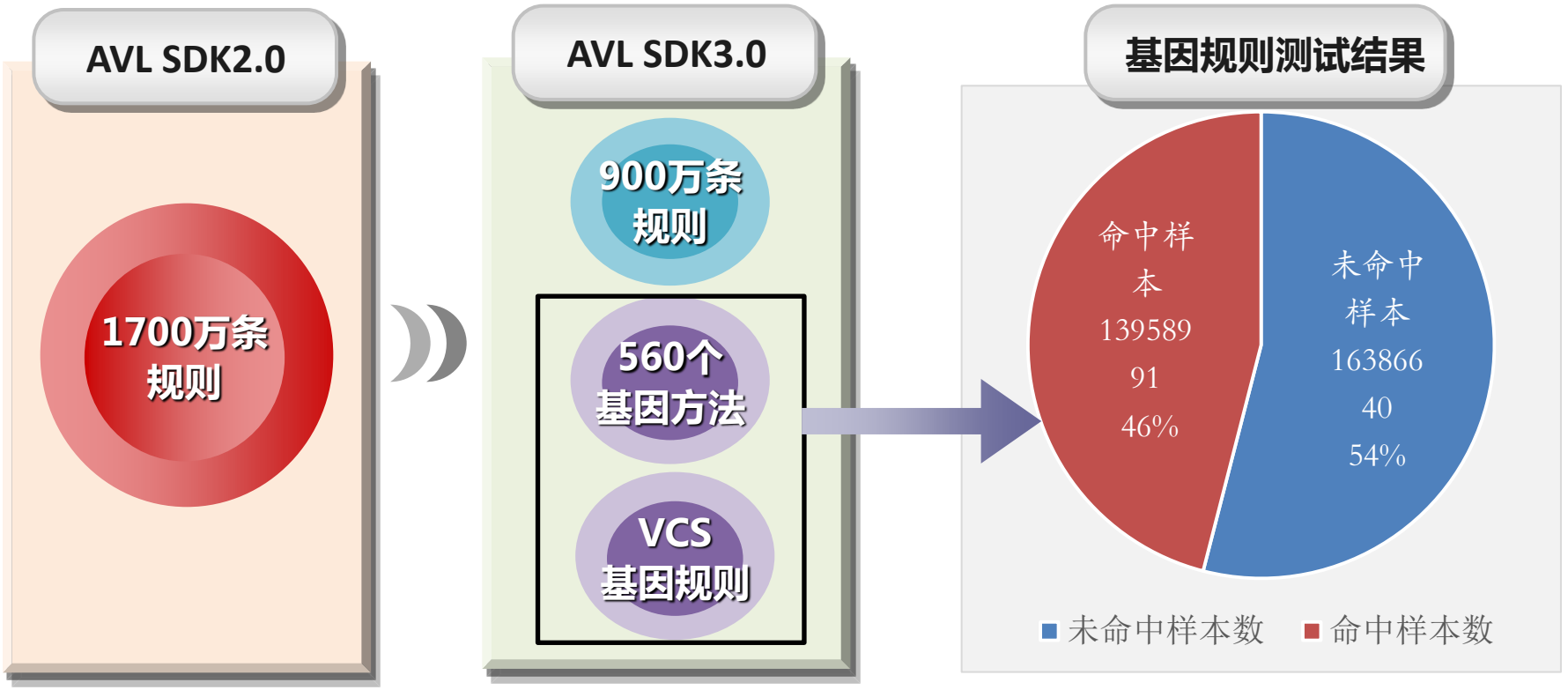
- 海量规则下的高速引擎

可定制

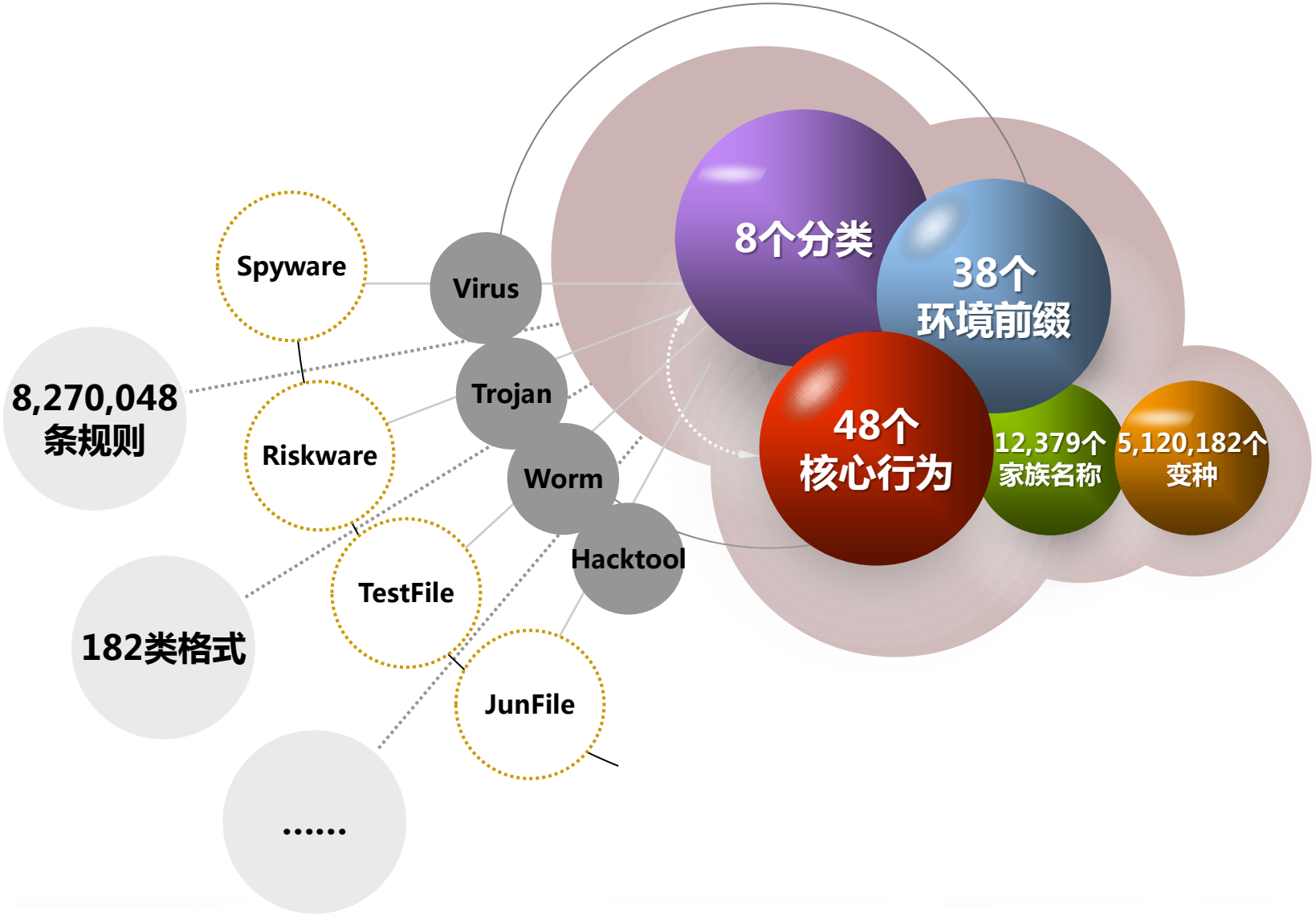
- 不同硬件平台、系统平台的支持（X86、ARM、MIPS）
- 不同资源需求支持（32M~G级别内存）
- 不同层次的检测支持（包、流、完整文件）
- 满足厂商对引擎、病毒库的个性化需求定制

我们今天还在做的-更精简有效

- 资源与检测效率平衡



高精度的检测能力

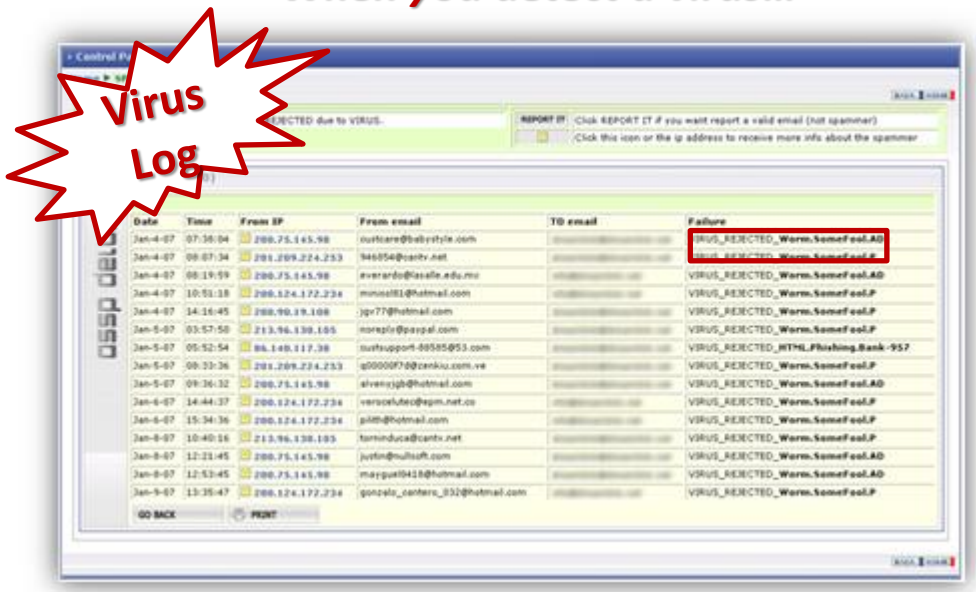

Spyware
Virus
8,270,048
条规则
Trojan
Riskware
Worm
Hacktool
TestFile
182类格式
JunFile
……
8个分类
38个
环境前缀
48个
核心行为
12,379个
家族名称
5,120,182个
变种

AVL SDK（DEMO SHOW）

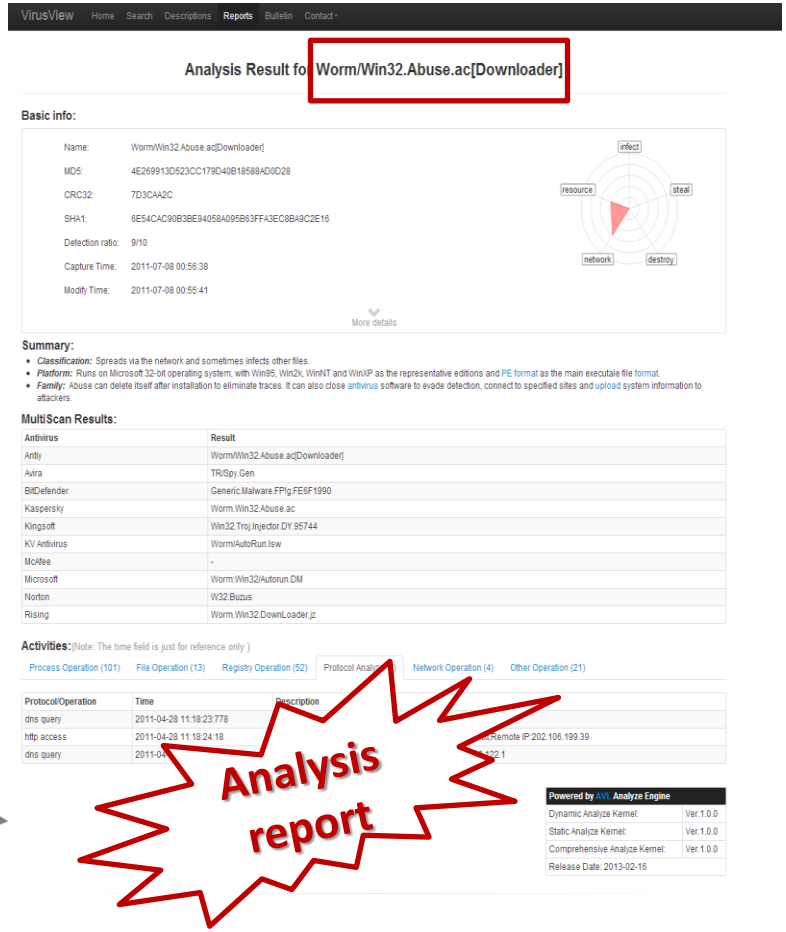
⊙演示安天下一代反病毒引擎的机理。

More infomation…

When you detect a virus…

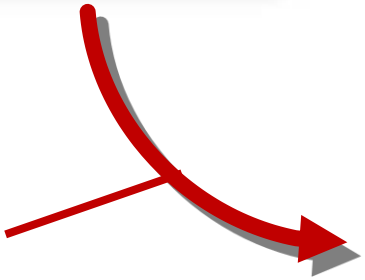


We also provide to you…



Get a Virus name

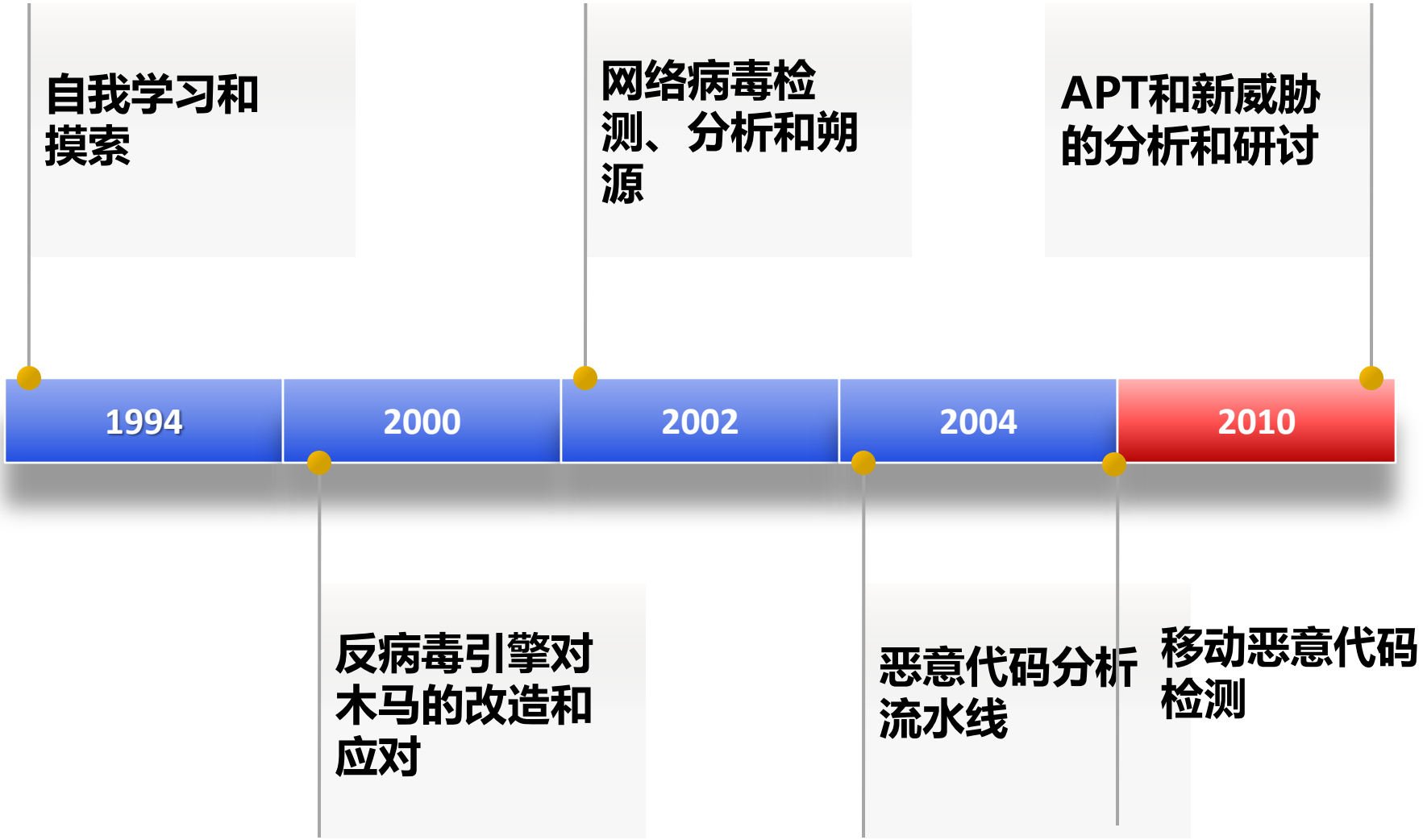
Worm/Win32.Abuse.ac



DEMO SHOW

⊙演示安天病毒百科 virusview.net

关于我的那些时间


自我学习和摸索
网络病毒检测、分析和朔源
APT和新威胁的分析和研讨
1994
2000
2002
2004
2010
反病毒引擎对木马的改造和应对
恶意代码分析流水线
移动恶意代码检测

关于我的那些语言

- 基于网络流和包的病毒检测（2002，xcon）
- 反病毒引擎的取证化应用（2003，xcon）
- 细粒度可嵌入的反病毒引擎（2004，xcon）
- 网络病毒监控系统的架构体系与研究方法（2005，Xcon）
- 后“冷战”时代的病毒捕获体制（2006，武汉大学）
- 反病毒引擎、产品和体系的安全挑战
- ……
- 管中窥豹——Stuxnet、Duqu和Flame的分析拾遗与反思（201，新安全威胁论坛）
- apt对传统反病毒技术的威胁(2012，CNCC)
- 从木马雪崩到APT的关联与必然（ 2012，XDEF ）
- 抽条的发动机——网络设备的第三方反病毒引擎反思（2012，ISF）
- 寻找apt的关键词——apt的本质思考（2013，ISC）
- 点射APT——高级持续性威胁分析的案例与方法思考（2013，中科院）
- 恶意代码对抗体系演进的四部曲（2013，国防科技大学）
- 走出蠕虫木马地带（2013，XDEF）

感谢同事们的努力

- 2001，传统引擎对木马的应对。
- 2002，千兆带宽的万级别规则检测。
- 2004,WINCE平台病毒检测。
- 2006，四级（内容、厂商、行为、位置）受信机制的rootkit检测
- ……
- 2012，2013，AVL for Mobile AVTEST检出率持续前列。

尾声

科尼利厄斯·瑞恩，在电影版《最长的一天》结尾加入了一句台词：“他死了，我病了，你迷路了，这就是该死的战争。”

不屈者不死，不屈者会满身伤痕；不屈者会走到终点，但要记住的是，我们不要迷失……

——2010年11月13日给同事们的邮件