



网络空间威胁对抗与防御技术研讨会
暨 第九届安天网络安全冬训营

亂雲飛渡

资源代价与安全算力

探海算沙

——流量检测能力要求变化与算力代价

安天 | 网络安全产品中心

CONTENTS

目 录

01

无所遁形需要的能力及其算力支撑代价

02

新能力方向期待更多算力支撑

03

新变化-基于国产化载体的流量检测

04

资源代价的一些思考



网络空间威胁对抗与防御技术研讨会
暨 第九届安天网络安全冬训营

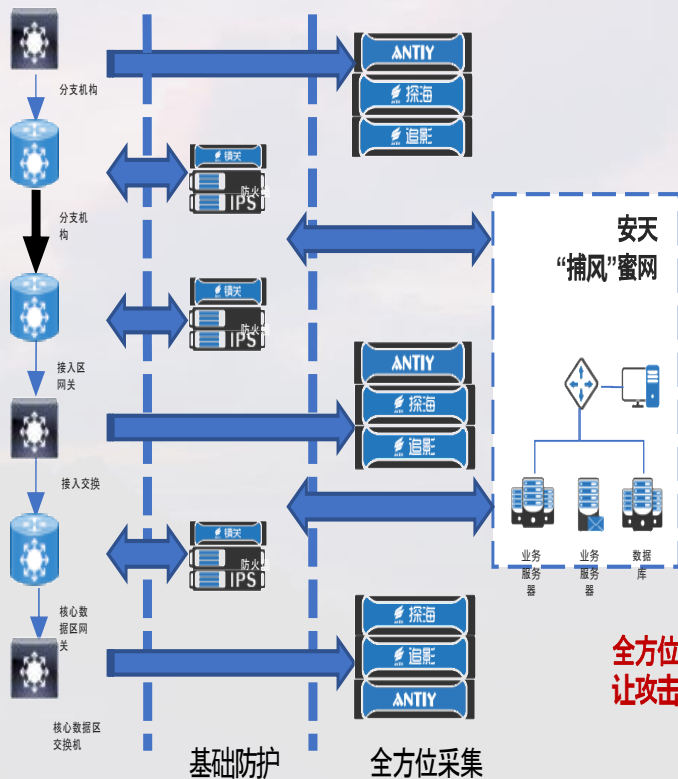
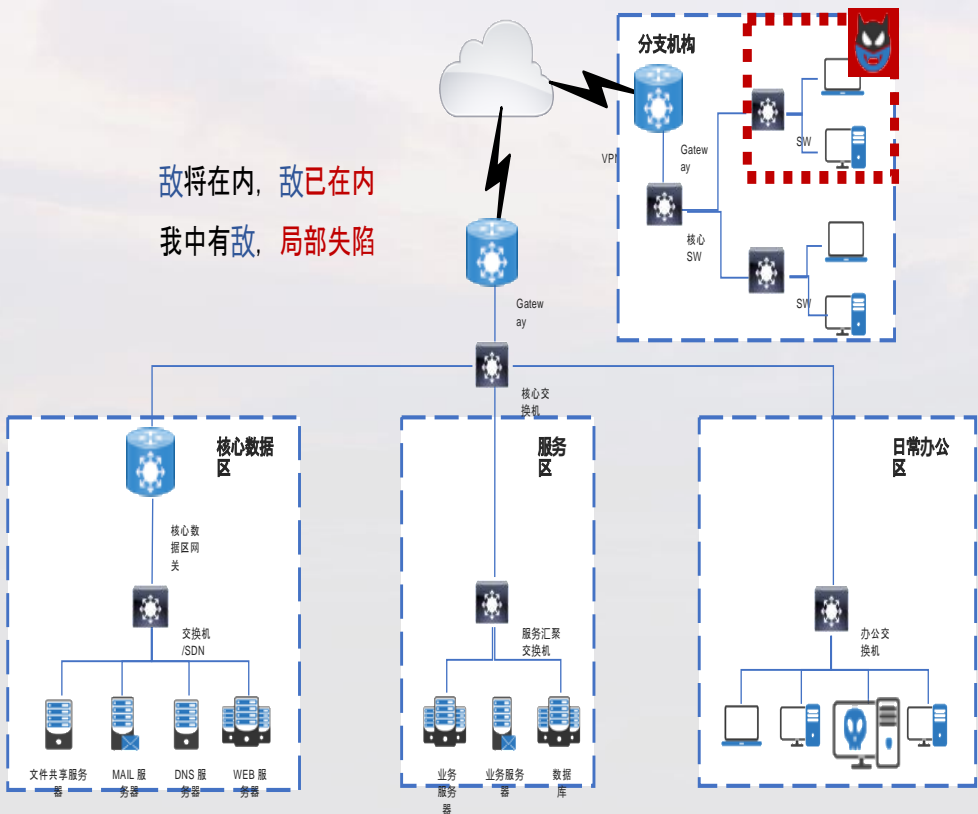
安天 | 智者安天下

01

无所遁形需要的能力 及其算力支撑代价

基于敌情想定，将检测能力部署在攻击者的必经之路

敌将在内，敌已在内
我中有敌，局部失陷



- 参考高价值目标构建合理分区
- 在抵达目标的路径上增加关隘
 - 业务路径
 - 数据路径
- 交叉火力覆盖无死角
 - 特别需要注意：设备管理流量
 - 特别需要注意：包头记录
- 被动防御能力是积极防御的基础

全方位采集、智能化响应
让攻击者无所遁行、无处可逃、无计可施

多维度的基础检测能力

全要素采集（丰富的检测对象）

多维度检测能力

威胁标注、追踪与响应

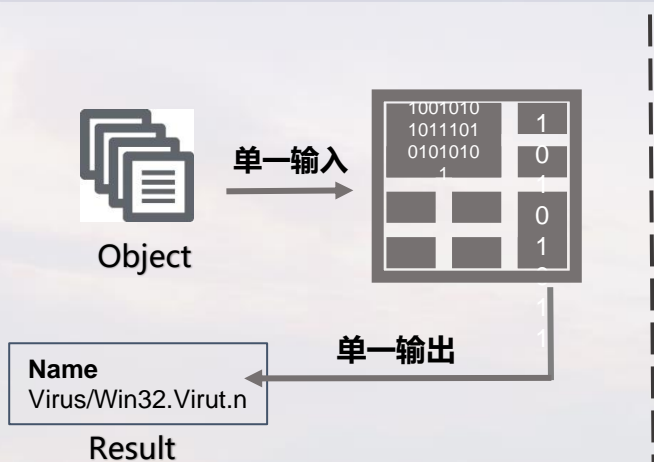


- 全量采集威胁要素，捕获高级威胁攻击中载荷高度定向、一次性投放；

- 获取载荷行为能力，形成对恶意代码的揭示能力，并对关键威胁信息进行留存；

- 通过威胁样本追溯威胁源和传播路径；
- 深度定制规则，提供威胁的向前追溯和向后守候能力；

威胁检测需要综合多个维度--多种输入输出对象



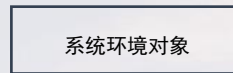
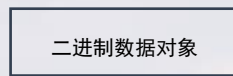
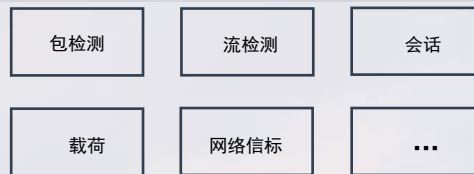
✓ 传统引擎

主要是以单一对象为输入，以单一结果为输出。而随着威胁的进一步演进和泛化，威胁检测已不能仅仅停留在对单一对象进行鉴定上。

✓ AVLSDK威胁检测引擎

多种输入对象，多种输出结果。威胁检测多样化。

网络层次检测



本地层次检测

多种输入

输出 1

- 黑白
 - 识别信息
 - 基础信息
- 多向量
 - 附加信息
 - 行为信息
- 核心行为
 - 远控 广告
 - DDOS 下载
 - 窃取
- 威胁行为
 - 传播 伪装
 - 隐蔽 对抗
 - 信息获取 攻击

输出 2

- 黑客组织名称
- 别名攻击目标
- 攻击领域
- 攻击方式
- 活跃时间
- 利用漏洞
- 组织简介

输出 3

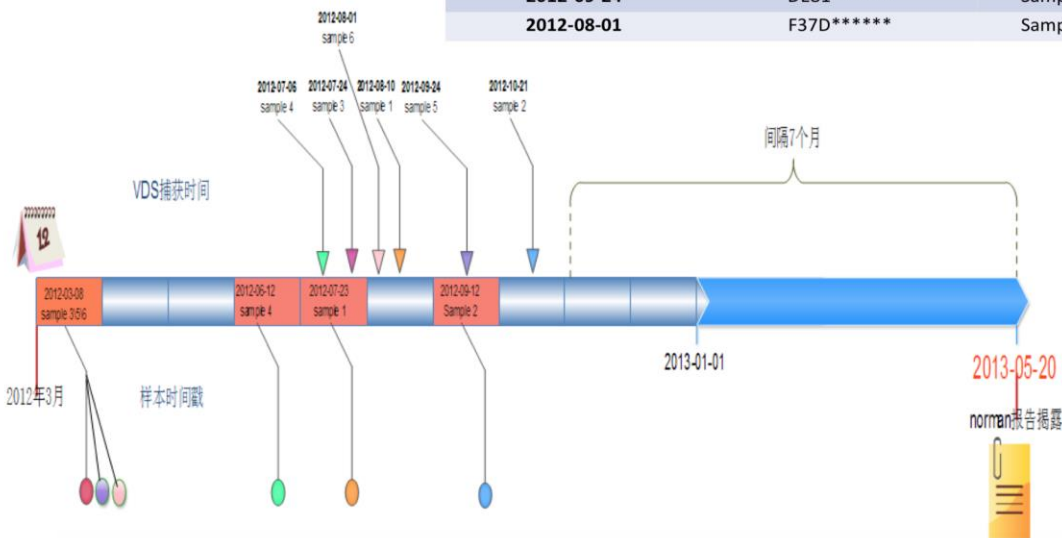
ATT&CK框架信息

初始访问、执行、持久化、提权、防御规避、凭证访问、发现、横向移动、收集、命令控制、渗透

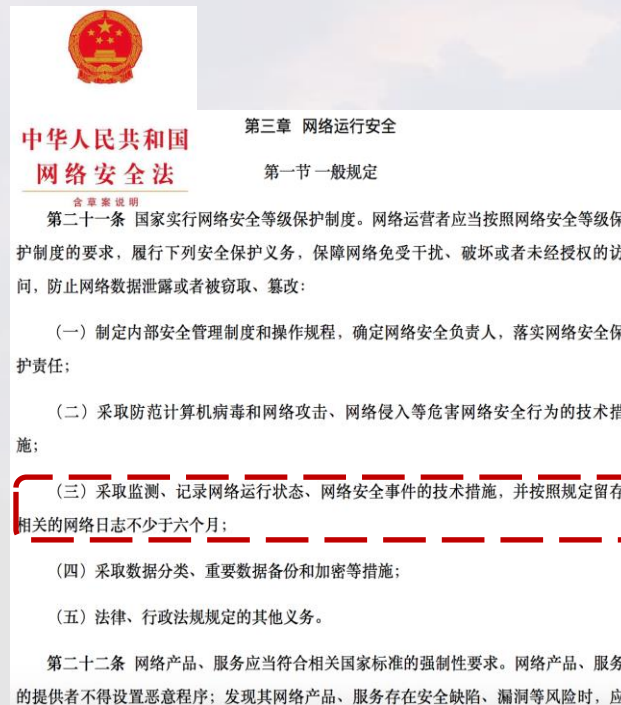
全要素长期留存，猎杀威胁，清点损失，复盘攻击

- HangOver对中国的攻击时间链——超过1年

捕获时间	样本hash列表	代号
2012-08-10	0D46*****	Sample 1
2012-10-21	734E*****	Sample 2
2012-07-24	9A20*****	Sample 3
2012-07-06	CE00*****	Sample 4
2012-09-24	DE81*****	Sample 5
2012-08-01	F37D*****	Sample 6



- 网络安全法要求日志保存不少于六个月



第三章 网络运行安全
第一节 一般规定

第二十一条 国家实行网络安全等级保护制度。网络运营者应当按照网络安全等级保护制度的要求，履行下列安全保护义务，保障网络免受干扰、破坏或者未经授权的访问，防止网络数据泄露或者被窃取、篡改：

(一) 制定内部安全管理制度和操作规程，确定网络安全负责人，落实网络安全保护责任；

(二) 采取防范计算机病毒和网络攻击、网络侵入等危害网络安全行为的技术措施；

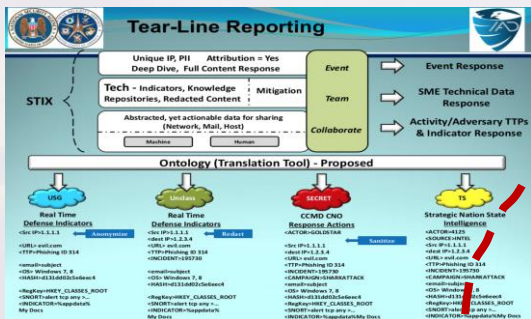
(三) 采取监测、记录网络运行状态、网络安全事件的技术措施，并按照规定留存相关的网络日志不少于六个月；

(四) 采取数据分类、重要数据备份和加密等措施；

(五) 法律、行政法规规定的其他义务。

第二十二条 网络产品、服务应当符合相关国家标准的强制性要求。网络产品、服务的提供者不得设置恶意程序；发现其网络产品、服务存在安全缺陷、漏洞等风险时，应

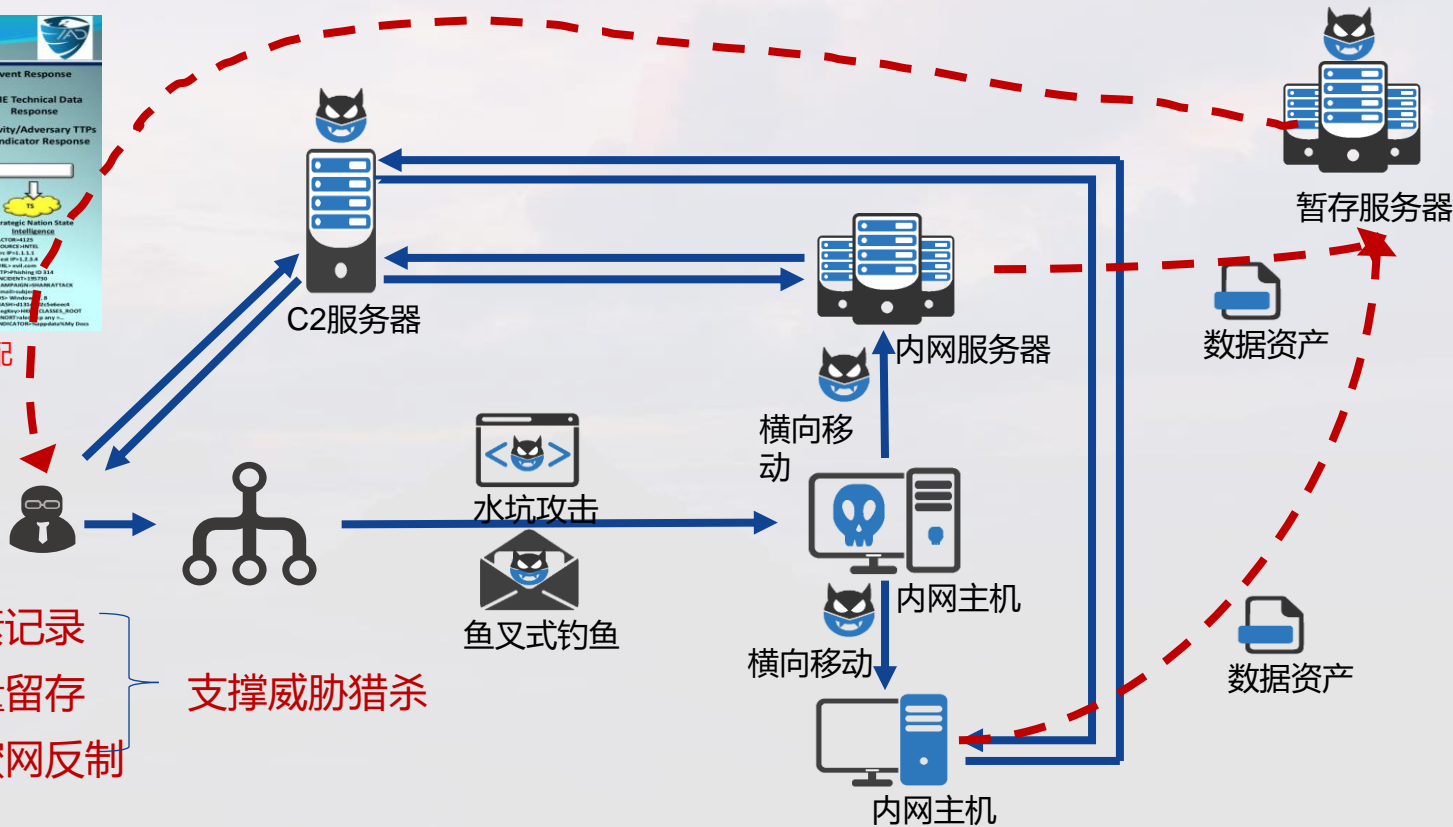
潜伏者被激活，内部威胁更需要全方位无死角的采集



基于场景，对威胁情报进行适配

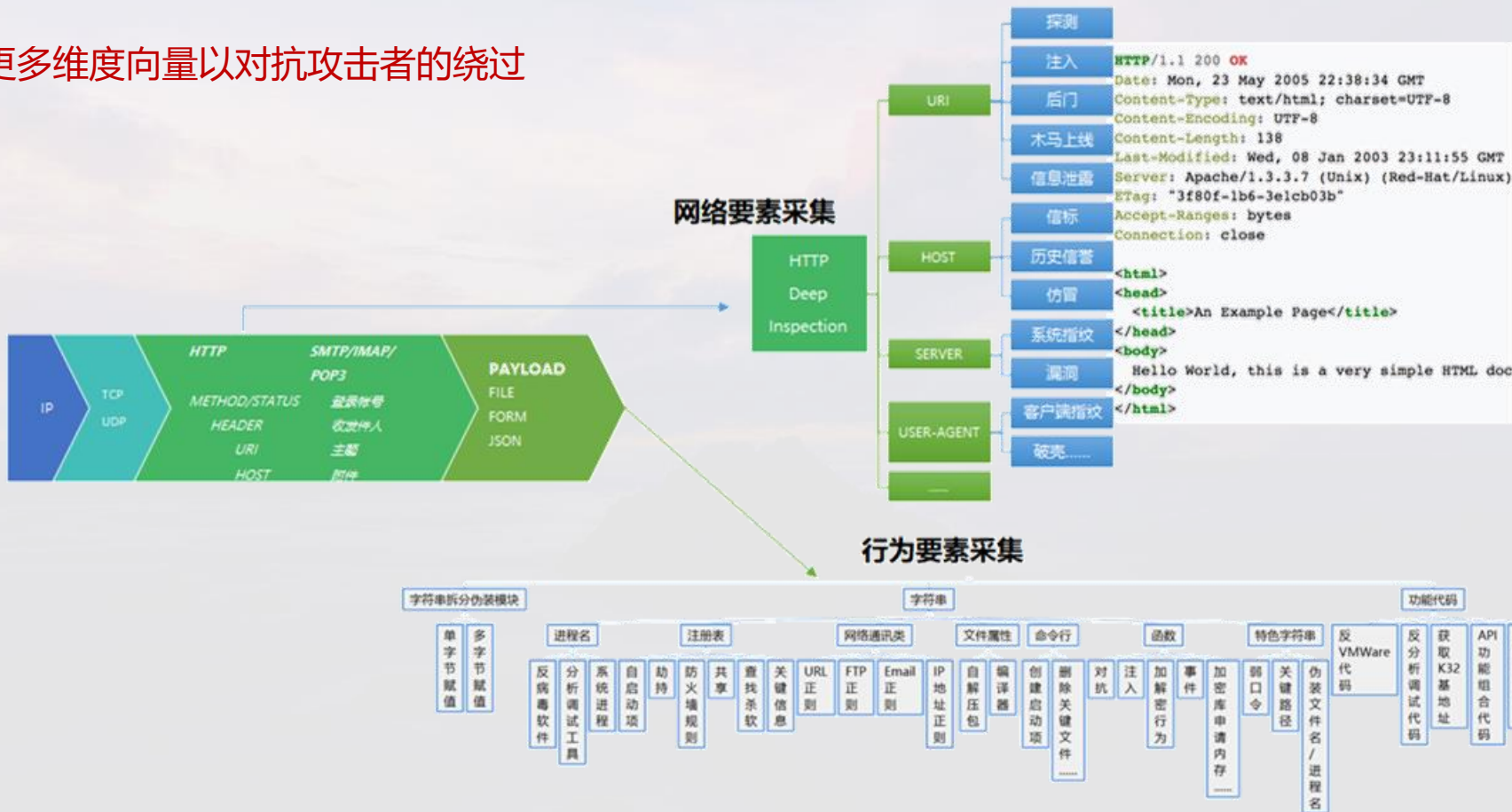
必经路径——全要素记录
 失陷节点——全流量留存
 信标触发——导入蜜网反制

支撑威胁猎杀



“探海”为融合威胁情报更丰富的要素采集

需要更多维度向量以对抗攻击者的绕过



有效应用向量级威胁情报，进行丰富有效的威胁检测

安天威胁情报系统基于引擎覆盖全球**100万台**网络设备和超**28亿部**智能终端的感知数据。以及持续对捕获样本的进行动静态分析，已构建**超百亿级别**威胁知识图谱。经过安天**20年**分析能力积累，持续输出生产包括机读情报和向量情报2类**20余种**威胁情报类型，以及标识超过**30种**威胁类型情报。在高级威胁分析场景提供完整的运营级情报的同源关联分析能力。在威胁检测场景向全系统供应向量级威胁情报。

安天是国内完整具备全域威胁感知、自主情报生产、高级威胁情报分析的全能力型情报厂商。

支持输入数据源

- 静态分析数据
- 动态分析数据
- 多引擎分析数据
- 流量探针数据
- 端点感知数据

- 高级威胁情报
- 自主生产情报
- 开源情报

支持输出的情报类型

机读情报

- 文件情报
- 域名情报
- IP情报
- URL情报
- 邮箱情报
- IP/域名端口情报

向量级情报

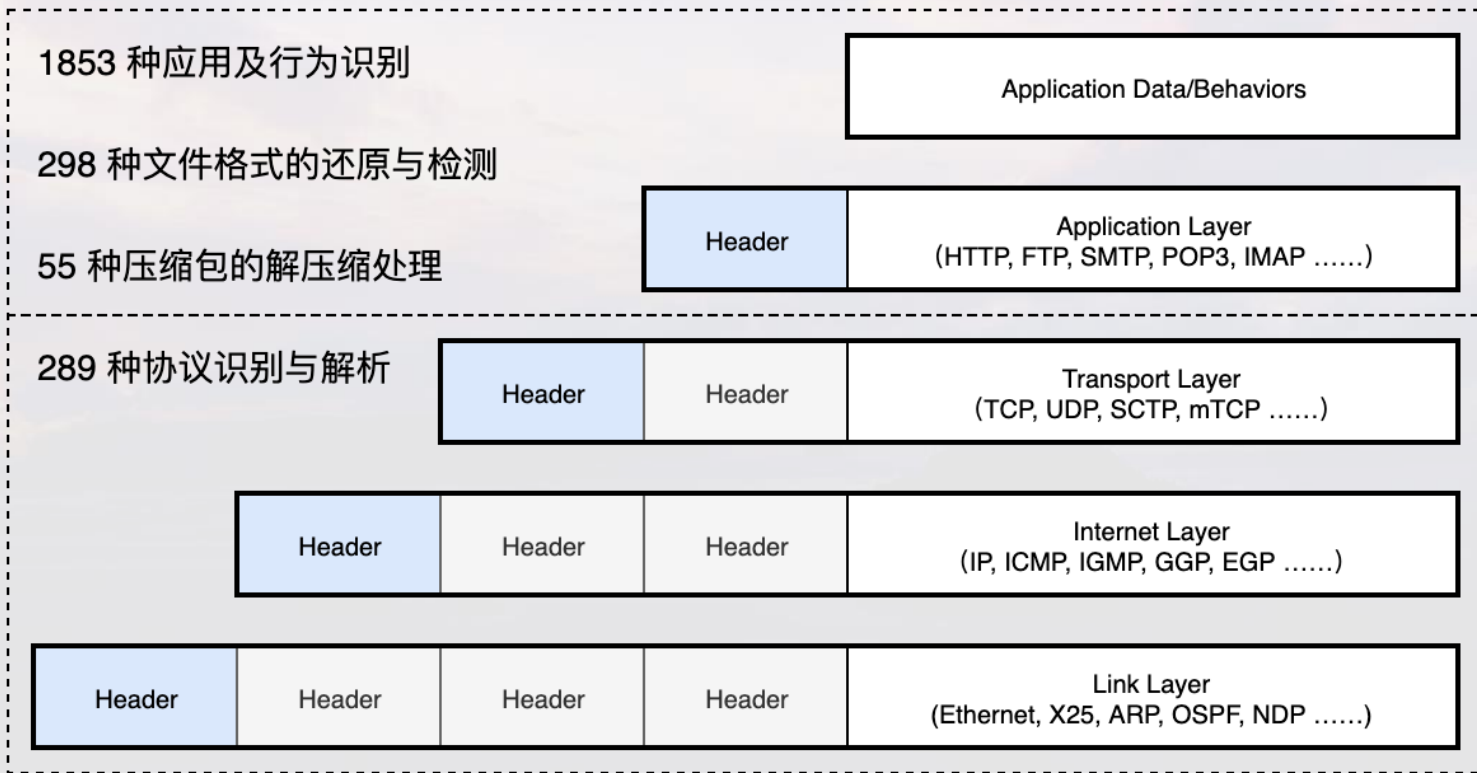
- **指令级向量情报**
- **API级向量情报**
- **功能级向量情报**
- **静态引擎输出情报**
- 注册表情报
- 互斥量情报
- 数字签名情报

威胁类型

- APT情报
- 僵尸网络情报
- C&C情报
- 勒索情报
- DDOS情报
- 挖矿情报

需安天下一代威胁检测引擎支撑

“探海”支持的协议识别、元数据化与要素提取能力



协议识别能力与开源项目对比

产品名称	协议/应用数量
探海	2142
OpenAppID	1464
nDPI	248
libprotoident	474

*所有开源项目基于 11 月 3 日版本统计

处理数据需要持续将经验转换为标签规则、场景规则

时间范围: 2016-12-01 00:00:00 到 2016-12-31 16:27:00 检索条件: 输入IP、域名、地区..... 搜索 高级 暂存

标签聚合: 跨境通讯, SMTP, 可执行程序

文件类型: PE .X86

目的IP(互联网): 188.166.95.178, 38.113.116.213, 173.194.67.27, 198.133.159.135

源IP(局域网): 192.168.18.162

概要描述 导出 标签聚合 IP 地理空间分布

时间	描述	备注/自定义
2016-12-23 05:20:08	局域网 192.168.18.162 通过 SMTP 协议 访问 美国 198.133.159.135 端口: 25 Spread Email 蠕虫程序 跨境通讯	
2016-12-23 01:58:47	局域网 192.168.18.162 通过 SMTP 协议 访问 美国 38.113.116.213 端口: 25 Spread Email 蠕虫程序 跨境通讯	
2016-12-23 01:57:10	局域网 192.168.18.162 通过 SMTP 协议 访问 荷兰 北荷兰省 阿姆斯特丹 188.166.95.178 端口: 25 Spread Email 蠕虫程序 跨境通讯	
2016-12-23 01:56:23	局域网 192.168.18.162 通过 SMTP 协议 访问 荷兰 北荷兰省 阿姆斯特丹 188.166.95.178 端口: 25 Spread Email 蠕虫程序 跨境通讯	
2016-12-22 19:13:50	局域网 192.168.18.162 通过 SMTP 协议 访问 欧盟 173.194.67.27 端口: 25 Spread Email 蠕虫程序 跨境通讯	

< 1 >

版权所有 安天 © 2016 { v6.6.0.2 }

行为向量 + 标签

多维度呈现

● 标签化

1. 减少用户需要关注的信息量
2. 传递标签背后的知识

● 场景化

1. 多个标签恰好构成了不同的场景
2. 自定义条件规则构成场景

【示例】识别特定攻击

跨境通讯、邮件通讯、压缩包、包含脚本

2016-12-23 05:20:02 开始
共计 9 个数据包, 738 字节

局域网 192.168.18.162 连接 局域网 198.133.159.135 25

邮件附件:

附件名称	类型	MIME	大小
locky	Archive(Robert_Jung.ARJ)	2985B37F7AC930211E0E200A8E72E8A0	24585

另一个例子：Log4J漏洞利用成功

局域网 192.168.10.46
通过 [HTTP 协议](#) 访问 (192.168.10.199:8000) 局域网 192.168.10.199 :8000 端口: 8000
发现可疑字节码文件: [敏感函数]疑似下载攻击组件-执行敏感函数
标签: [可疑字节码文件](#) [包含敏感函数](#) [RunTime](#)

ATT&CK™ [横向运动](#)
NSA | CSS [接触目标与进攻突防](#)

局域网 192.168.10.46
通过 [HTTP 协议](#) 访问 (192.168.10.199:8000) 局域网 192.168.10.199 :8000 端口: 8000
发现可疑字节码文件: [敏感函数]疑似下载攻击组件-执行敏感函数
标签: [可疑字节码文件](#) [包含敏感函数](#) [RunTime](#)

ATT&CK™ [横向运动](#)
NSA | CSS [接触目标与进攻突防](#)

局域网 192.168.10.46
通过 [HTTP 协议](#) 访问 (192.168.10.199:8000) 局域网 192.168.10.199 :8000 端口: 8000
发现可疑字节码文件: [敏感函数]疑似下载攻击组件-执行敏感函数
标签: [可疑字节码文件](#) [包含敏感函数](#) [RunTime](#)

ATT&CK™ [横向运动](#)
NSA | CSS [接触目标与进攻突防](#)

局域网 192.168.10.46
通过 [HTTP 协议](#) 访问 (192.168.10.199:8000) 局域网 192.168.10.199 :8000 端口: 8000
发现可疑字节码文件: [敏感函数]疑似下载攻击组件-执行敏感函数
标签: [可疑字节码文件](#) [包含敏感函数](#) [RunTime](#)

ATT&CK™ [横向运动](#)
NSA | CSS [接触目标与进攻突防](#)

局域网 192.168.10.46
通过 [HTTP 协议](#) 访问 (192.168.10.199:8000) 局域网 192.168.10.199 :8000 端口: 8000
发现可疑字节码文件: [敏感函数]疑似下载攻击组件-执行敏感函数
标签: [可疑字节码文件](#) [包含敏感函数](#) [RunTime](#)

ATT&CK™ [横向运动](#)
NSA | CSS [接触目标与进攻突防](#)

局域网 192.168.10.46
通过 [HTTP 协议](#) 访问 (192.168.10.199:8000) 局域网 192.168.10.199 :8000 端口: 8000
发现可疑字节码文件: [敏感函数]疑似下载攻击组件-执行敏感函数
标签: [可疑字节码文件](#) [包含敏感函数](#) [RunTime](#)

ATT&CK™ [横向运动](#)
NSA | CSS [接触目标与进攻突防](#)

局域网 192.168.10.46
通过 [HTTP 协议](#) 访问 (192.168.10.199:8000) 局域网 192.168.10.199 :8000 端口: 8000
发现可疑字节码文件: [敏感函数]疑似下载攻击组件-执行敏感函数
标签: [可疑字节码文件](#) [包含敏感函数](#) [RunTime](#)

ATT&CK™ [横向运动](#)
NSA | CSS [接触目标与进攻突防](#)

可执行程序 25 种格式, 已选择其中 25 种

- LE
- IOS
- EPOC
- MAGIC
- MENUET
- ODEX
- CIGAM
- NDS
- app
- CLASS
- PALM
- DOS

2022-01-12 21:27:49 开始
共计发送 6 个数据包, 522 字节

局域网 192.168.10.46 :52299
连接 (192.168.10.17) 局域网 192.168.10.17 :80

GET 请求:

HOST	192.168.10.17
URI	/Exploit.class
User-Agent	Java1.8.0_181
Host	192.168.10.17
Accept	text/html, image/gif, image/jpeg, */*; q=.2
Connection	Keep-Alive

通过识别class文件传输

探海可直接发现Log4j漏洞利用成功事件

基础数据标签化、场景化，构建定制化威胁守候能力

编辑追踪条件

符合以下场景时

邮件通讯

检索条件: 192.168.16.24*

跨境通讯

APT-TOCS[2]

疑似“海莲花”的一些信息。

将下列对象

信标 互斥量

标注为

LM 杀伤链模型 武器构建/商业军火

标注为

- APT 生存周期模型
- 侦查探测/信息收集
- Search
- APT 生存周期模型 ✓
- LM 杀伤链模型
- 自定义标签

编辑追踪条件

符合以下场景时

信息泄露

检索条件: 192.168.16.24*

跨境通讯

APT-TOCS[2]

疑似“海莲花”的一些信息。

将下列对象

事件

标注为

APT 生存周期模型 侦查探测/信息收集

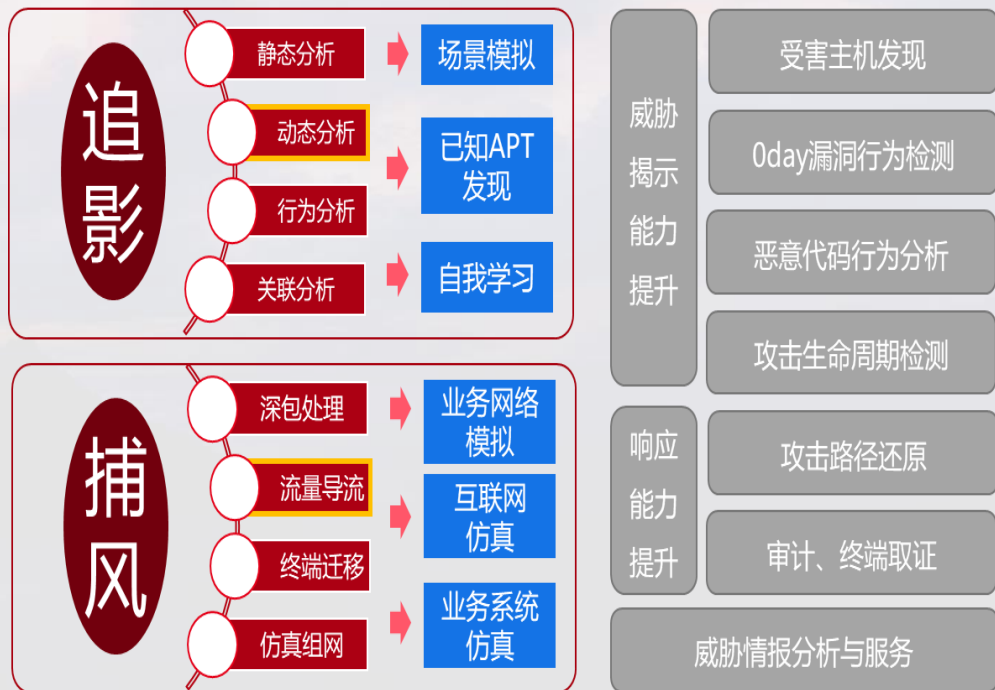
放弃 保存

全面掌握资产、实现实体分析需要全要素支撑

• 全面掌握资产以支撑场景化的分析



• 构建基于我情的沙箱与蜜网



探海：具备骨干网30Gbps全流量检测能力的ND20000型号



- 基于X86纯软件实现，2U设备载体
- NUMA高效算力使用
- RSS、FlowDirector等网卡功能应用，加速处理
- 30Gbps骨干网流量全流量高速匹配、文件还原能力
- DDoS等网络攻击检测过滤
- 同时有效跟踪2000万连接
- 超大文件还原，支持跨会话分片还原
- 支持100万以上浮动关键字；1万以上正则表达式匹配
- 静默功耗：300W；
- 全负载有效功耗：450W；
- 最高功耗：600W

检测分析	场景关联检测		文件特征检测		行为向量检测
	单包检测	单向流量顺序检测	C&C通联检测	自定义检测	SNORT规则导入
协议识别解析	流式协议元数据分向解析			文件解析	
	单包协议猜测	状态转换协议特征		基于协议特征的乱序重定位	基于文件特征的传输识别
	包暂存	有限状态机	数据包保序		
流量接入	数据包池	跨进程通讯	零拷贝内存访问		
	非一致内存访问		直接内存访问	大内存页	内存映射
	分光流量数据接入			TAP分流流量接入	

骨干网30Gbps全流量检测能力的算力代价

CPU: 2颗 16核至强处理器, 开启超线程
内存: 384G 硬盘: 4T*2 (RAID1)
网络外发: 全要素日志外发占用约200MB每秒 (1.6Gbps)



报文捕获: 2C
TCP流还原检测: 10C
UDP流还原检测: 5C
DNS流还原检测: 1C
元数据化、补充上下文信息: 4C
标签场景规则检测: 6C
模型检测: 6C
数据外发: 1C
分析索引建立: 5C



数据报文池: 64GB
(最大保留4192万个报文)
简易追踪流表: 9GB
(约1亿项, 每项96字节)
深度解析流表: 60GB
(2048万项, 每项3K)
检测模型: 20GB
内部消息通讯内存池: 12G
元数据化: 30GB
基于统计的分析: 80G
分析索引、缓存: 105GB



检测模型读取: 35iops (50MBps)	心跳日志发送: 0.01Mbps
索引写入、还原 文件写入: 10 iops (20MBps)	威胁日志发送: 1Mbps
查询分析: 120iops (300MBps)	元数据化的全要素日志外发: 每秒200MB (约20万eps)

探海：海量流量接入处理上的一些经验

以空间换时间

预分配内存池
每个内存分片大小的确定
分段内存
JumboFrame处理

充分利用硬件

Receive Side Scaling
多队列跨CPU分配
VxLAN、MPLS等云环境的哈希不均衡问题解决
FlowDirector的32K表利用



高效

内存直接访问DMA技术
NUMA
Kernel By Pass
HugePage



流式处理

状态迁移方式识别解析
跨内存分片处理匹配
引用计数暂存报文



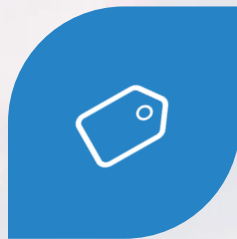
超量流量应对

Syn、NTP、ICMP、DNS的DDoS
报文伪装、反射
受害者识别&流量过滤
简易流表 / 深度处理流表

探海的元数据化及其算力消耗

解码

正常HTTP传输的压缩：Gzip/Deflate
MIME编码：Base64、Quote-Printable
URL编码及其利用的规避检测
数字证书：ASN.1
域名Punycode编码及其规避检测
消耗CPU
尽量利用SIMD指令，仍有更高性能期望



上下文配对

协议头跨包解析留存
HTTP协议头，邮件传输协议头
请求-响应双向在同一会话配对分析
Webmail、FTP等协议跨流配对分析

消耗内存
尽量利用零拷贝能力

识别

基于协议内容特征的协议识别
可识别443端口传递非加密流量
乱序/未成流流量，根据内容恢复协议
标签化、标签规则、地理位置信息库



编码发送、索引检索

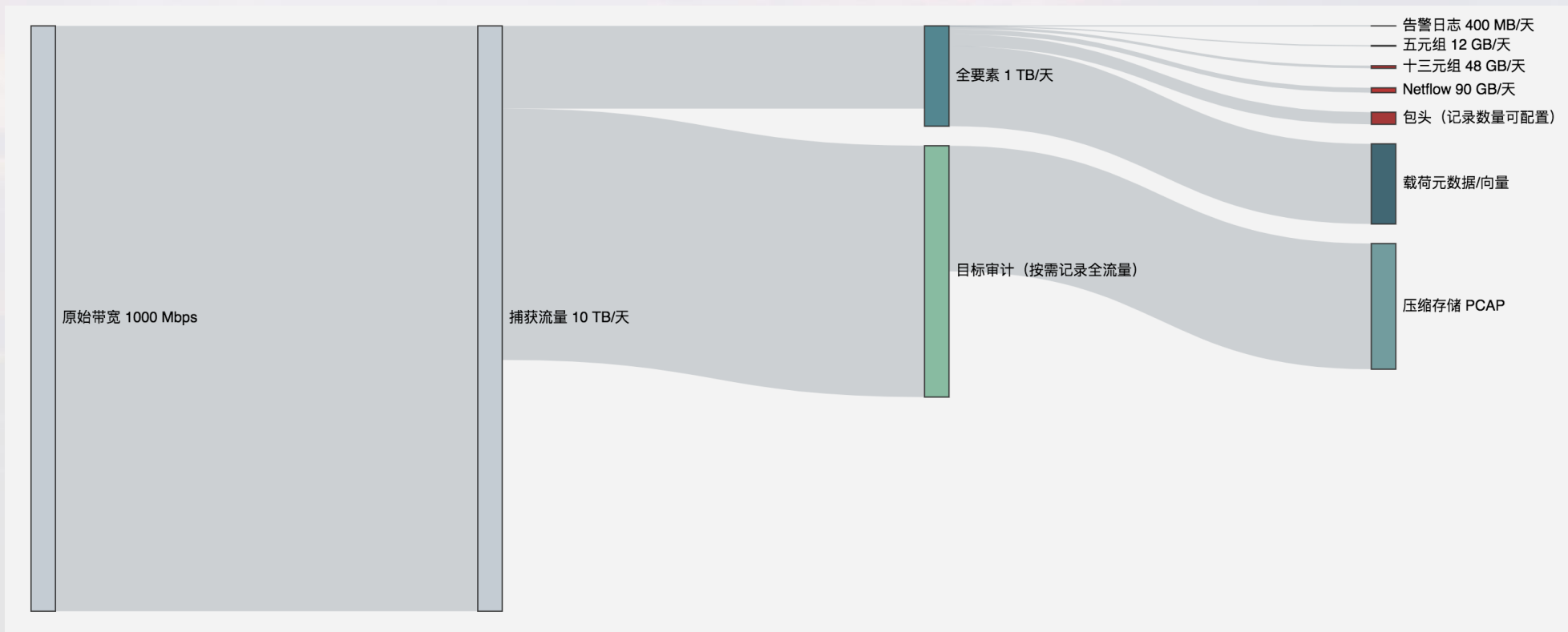
编码、解码：20万EPS
消耗CPU

发送：消耗网络带宽

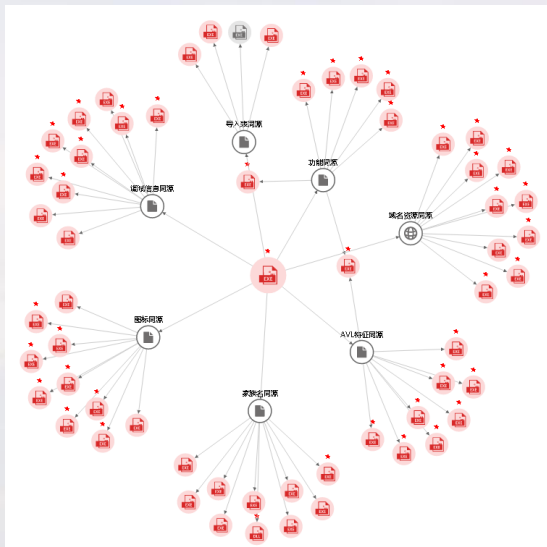
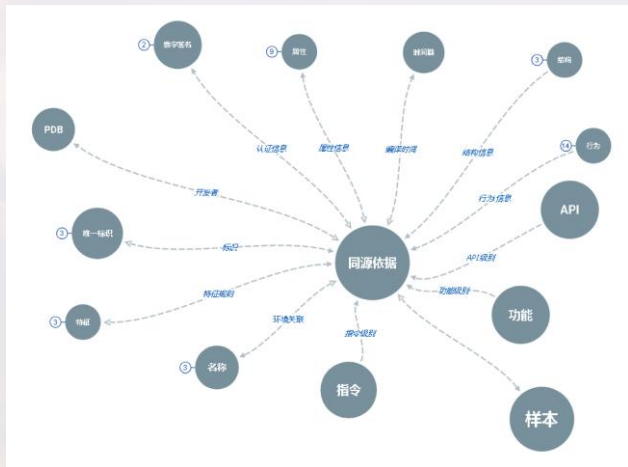
索引：分词，前缀，倒排索引
消耗IO+CPU

消耗CPU

海量数据带来的数据留存分析挑战



更多的威胁检测能力期待更多算力支持



- imphash 库
- ssdeep 库
- Vhash 库
- Authentihash 库
- Rich PE header hash 库
- TLSH 库
- Section Hash 库

通过静态分析引擎对样本分析生产输出的向量提取点，依据同源分析维度和专家研判，形成向量级威胁情报。在自动化分析和人工分析场景提供**抗变性极强**的检测能力。

基于安天静态分析引擎的向量级威胁情报，将**APT白象**组织使用的攻击工具聚合在一起。但这需要大量算力支持。

有效应用威胁情报所需安全检测所需的特殊Hash算法

但待匹配数据的准备和匹配所需算力依然匮乏



网络空间威胁对抗与防御技术研讨会
暨 第九届安天网络安全冬训营

安天 | 智者安天下

02 新能力期待更多算力支撑

NDR的呼唤：情报驱动，AI助力的高级威胁检测

看得见

想得清楚

说的明白

可处置



规则覆盖和优化、威胁情报、智能化算法

攻击确认、攻击链、证据链、ATT&CK

事件还原、场景分析、扩线分析

拦截封禁、断网隔离、溯源反制

基于专家经验的数据处理逻辑思路，基于流量检测数据进行智能化推导分析，解决在不同场景下的业务需求

看得见：基于文件行为序列泛化的向量检测能力

✓ BinExecute/Microsoft.PE

a) 静态向量

b) 远控静态配置解密

• 行为标签 (共162项):

对抗, 传播, 控制
隐藏, 窃取, 欺骗

• API (51类):

模块相关, 网络相关
文件相关, 进程相关
窗口相关, 内存相关

- IP, URL
- MAIL, DOMAIN

- 证书信息: 颁发者, 使用者, 有效期, 算法...
- 签名信息: 证书链, 签名人名字, 签名时间
- 判定标签: 伪造, 吊销, 过期, 证书不完整

✓ OFFICE

- 文档说明: 标题, 主题, 标记, 类别, 备注...
- 文档来源: 作者名, 公司, 版本号, 管理者, 创建内容时间...
- 文档内容: 夹带文件, 宏代码OLE对象...

向量提取

相似计算

行为划分

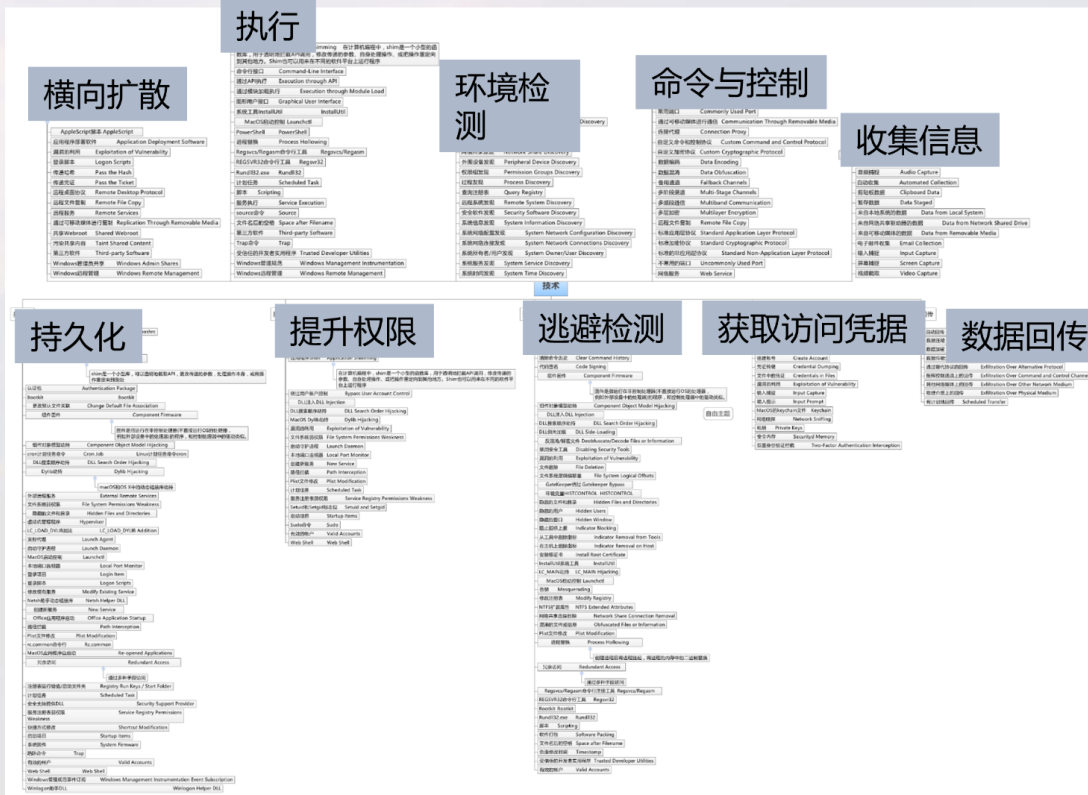
误报过滤

告警

CPU密集型

主要算力消耗:

提取 (查找树遍历)、相似度计算



看得见：语义分析及污点传播执行以应对webshell



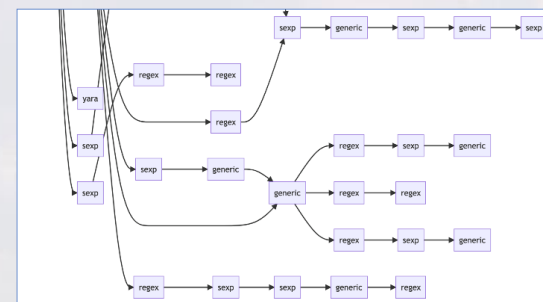
典型家族 100% 检出

- Godzilla
- 冰蝎 2/3
- P.A.S Webshell
[APT: Sandworm Team]
- Caterpillar
[APT: Volatile Cedar]
- AA21-259A
[APT: 未知组织]
- ChamelGang

基于知识图的检测逻辑

344 个节点构成的有向无环图

字符特征 (单词、常量和变量名等)
语法结构 (函数调用、对象创建和赋值语句等)
语义 (先创建再调用、循环处理字符串、删除文件)
——进一步分析污点路径 (输入数据变代码、变文件路径)



CPU密集型

主要算力消耗:

反混淆。

语法树构建

语法树推导、语义分析

看得见：自适应webshell加密通讯应对



- 对威胁的理解比随意套用模型算法更重要
- 基础数据的准备很重要——刨除协议本身的压缩和加密
- 从特定的事情做起——webshell加密流量应对
 - 与图片区分；与二进制通讯流量（protobuf等）区分；
 - 与特殊通讯协议区分（mmtls）




$$H(X) = - \sum_x P(x) \log_2 [P(x)]$$

香农信息熵用于度量通讯中的信息量

CPU密集型：
对目标路径的每个包计算概率
遍历、自增——每个包的每个字节
求概率——每字符
遍历、求和、log

```
Cookie: PHPSESSID=jd1q99b814bs148snhhu1h0jn;
3Mh1yMtoZViV5wotQHPJtwwj0F4b21yToNK7LfdUn7zmyQFfx/
za1GulHkg+85lclgrXRNeC6Qwv0n1NpheyYv15A1JgttA4T4g4VHzDhp3FYoo8AFL11D1Jq
Ph9dSSD3kuyn5Kg8NtU376ivmavtswakm3HBP8mgnQDg1ByMuMk0ek6fat7jpbunAJwCJFZ
Y9KP/5rE6gpaHfIy0Shv1PScvr3E1oCoLuvaf+hOr+/4q+UFzHeLcw1fH68hQ28J/
S3yDfQhMBrN0g8tNkx7F6p70o06WfPqna+9C3DA60yXuoerPxEewq3ppKk7k81ZFmm8x001e
ycZMgl36RSgm00Jbu80z+v4ZpCfTvch755An0i1tMbxZ9GC94/
g2rYnNc9x4vQc3urYgNKUMxhbkouN2g8WxL2H31vEDgrEkLSM/
JfBTj8g140X7IhCr80MAGAc/+xHk25s0oKQ0B15/
gCgcn3coQhMxMzvn0FDhtuYwKQBYfL0/
f1RALvPnEEZm4qZ1vIUvoaYKm2PYOANGHrQ8vxvNL4FMDMHj5XbQ9/e0/
vFNCChZvz+1o0TKdo8067NgogxBJ05P/
qwa6dakPWYFR4KMrUq1qnakmk+cG5zRup616Bd8E39tb8WmsqDAMX101TY2Ql04BFZbHA4
SKwaTmnW3aVUZ0DeBwFKNj8B/QTqaGMBPC0A41uv9TV56XKoD5L6t5//jrnlyZ7nXYU/
x309k7wdd1YcAZgk5dzMrAsR7nidYYt6IX8U0U21X2IWM65ISZaho8Xrs+MsqaxxGRT/
GGeqrEwXofDge04kQEaqod7WAfPsqE11F1Ee1gh9F0JHGHP43b8q0BmXWeidXxk55sq313
hiJTS7INnIyRztS91DShdFy6Y2rnhEYojw5hcu6GZivTpJLGj535/
X69zqnmurD64e2A4o6iHeItcoj/1p6WU1z2DCndu/9DohnYZ0c/
IJ5xgY8BmzYsg4M6KLrdqUtRMC7ko3Yqurfxuv1VhiULHWHIF87K9RhtB3+er+kMKL2/
bPWOC1cYpngRXdo0dK4jY171CS+fgx+RQKpCYoij6kq36PUUVEGZwLUdRtE4DvOrXhhT
00000090 53 69 74 65 3d 4c 61 78 0d 0a 55 73 65 72 2d 41 Site=Lax..User-A
00000090 67 65 6e 7a 3a 20 4d 6f 7a 69 6c 6c 61 2f 35 2e gent: Mo zilla/5.
00000080 30 20 28 57 69 6e 6a 6f 77 73 20 4e 54 20 36 2e 0 (Windos NT 6.
000000C0 31 3b 20 57 4f 57 36 34 29 20 41 70 70 6c 65 57 1; W0M64 ) AppleE
000000D0 65 62 4b 69 74 2f 35 33 34 2e 35 30 20 28 4b 48 ebKt/53 4.50 (KH
000000E0 54 4d 4c 2c 20 6c 69 6b 65 20 47 65 63 6b 6f 29 TML, lik e Gecko)
000000F0 20 56 65 72 73 69 6f 6e 2f 35 2e 31 20 53 61 66 Version /5.1 Saf
00000100 61 72 69 2f 35 33 34 2e 35 30 0d 0a 43 61 63 68 ari/534. 50..Cach
00000110 65 2d 43 6f 6e 74 72 6f 6c 3a 20 6e 6f 2d 63 61 e-Contro 1: no-ca
00000120 63 68 65 0d 0a 50 72 61 67 6d 61 3a 20 6e 6f 2d che..Pra gma: no-
00000130 63 61 63 68 65 0d 0a 48 6f 73 74 3a 20 31 30 2e no-cache..H ost: 10.
00000140 32 35 30 2e 32 35 2e 33 36 0d 0a 41 63 63 65 70 250.25.3 6..Accep
00000150 74 3a 20 74 65 78 74 2f 68 74 6d 6c 2c 20 69 6d text/html, im
00000160 61 67 65 2f 67 69 6e 2f 69 6d 61 6f 65 2f 6a age/gif, image/js
00000170 70 65 67 2c 20 2a 3b 20 71 3d 2c 2c 20 2a 2f *; q=2, /
00000180 2a 3b 20 71 3d 2c 3d 0d 0a 43 6f 6e 65 63 74 /*; q=2, /
00000190 69 6f 6e 3a 20 6b 65 65 70 2d 61 6c 69 76 65 0d ion: kee p-Alive.
000001A0 0a 43 6f 6e 74 65 6e 74 2d 4c 65 6e 67 74 68 3a .Content -Length:
000001B0 20 38 32 32 34 0d 0a 0d 8224...
00000189 9b 1c 86 25 ba b6 27 f9 0b 83 c4 ee 10 85 04 14 ..%.'. ....
000001C9 86 26 18 c1 20 37 54 5a a6 94 ab 44 04 2a e6 12 &.. 7TT ...D.$..
000001D9 96 b6 4e 08 4c 76 49 e3 22 a1 af 30 dc 24 f8 71 ..N.LVi. "...0.$..q
000001E9 e4 64 f2 dd a7 23 e8 2c e3 19 0e c1 cb ff 0a a6 .d..#, ..
000001F9 bf 34 55 80 87 f5 05 f5 38 c2 0f 67 a8 6a 44 4f .4U..... 8..g..J0.
00000209 fc 74 c2 70 a1 38 6a a8 ff 1c 21 08 11 b6 35 d4 .t.p.8j. ...1..5D.
00000219 7f 0b 92 8a 7c 78 56 5d 6f 67 8d ea c9 34 13 fd ...].V. ox...4..
00000229 f2 23 41 d0 28 c4 78 d9 b8 08 7b 3c 03 ee 09 08 .#.A.(x...{....
00000239 21 cd 37 3a 39 af 39 11 66 75 fd f3 4a 1a b2 f0 !.7:9.9. fu..J..
00000249 0a 62 93 9e 09 ea 55 03 07 18 ff 82 73 af d0 .b....U ....s...
00000259 b7 cc b6 a0 d6 49 d2 ee 33 d1 20 cc df f4 a9 19 ....I.. 3. ....
00000269 b7 8b 51 b8 0b 4d a9 aa 79 58 c7 fe 43 43 e7 e6 ..Q..M.. yX..CC..
00000279 73 43 14 65 f4 00 99 85 33 9b b6 b7 95 35 82 0b sC.e... 3...5..
00000289 cb 91 c1 55 df b0 ce 88 3d 41 be d7 41 25 4f 5d ..U.... =A..AX0]
00000299 55 35 6b 96 ae 21 f4 ac af 45 85 51 46 f1 59 b6 USk..!.. .E.QF.Y.
000002A9 24 38 9a dc 6c 27 83 7d 5c 8e a5 fd 88 0d 91 e7 $8..1'.') \.....
000002B9 c7 6b 0d 6b 99 49 0f ee f6 23 ff 21 14 0c 82 56 .k.k.I.. #.1...V
000002C9 c3 c5 aa 65 46 4b 2d 94 97 21 76 61 4b 5f 12 5b ...eFK...lvaK_[
000002D9 b1 97 35 79 4c 1b 24 9c 8b 2c 51 2d 53 2e 35 23 ..5YL.$...Q-5.5#
000002E9 fe .. 09 37 .. 09 37 14 .. 6c 63 1b 5b 20 8d .. 27 ..
```

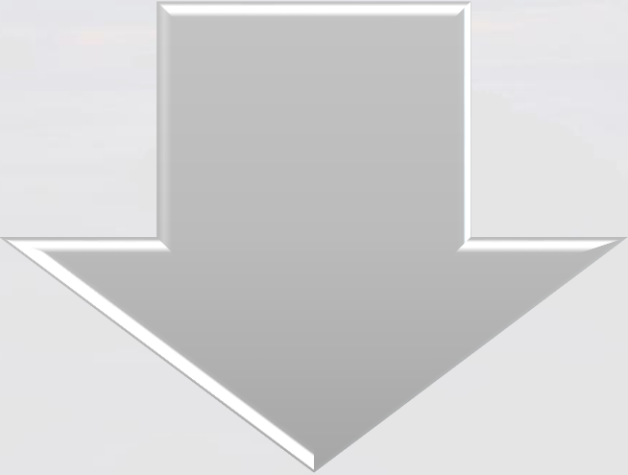
看得见的未来？基于机器学习的威胁检测建模？



- 使用随机森林算法解决PE类型恶意样本检测
- 使用随机森林算法解决PDF类型恶意样本检测
- 使用关联算法实现攻击链自动关联取证
- 使用机器学习解决恶意安卓应用检测
- 使用深度学习解决恶意二进制样本检测
- 使用机器学习实现用户网络行为分析
- 使用人工智能解决异常/加密网络流量检测
- 使用相似性算法实现恶意软件同源性分析
- 使用回归神经网络检测二进制恶意软件

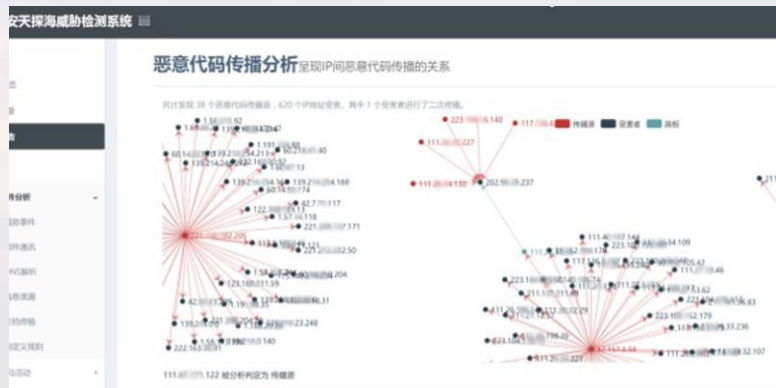
更强大，更智能的威胁检测能力
需要更强大的算力支持。

能力、效果和性价比的平衡如何达成



- 模型大小——动辄上百GB
- 原始数据清洗——准备对应向量需要大量CPU
- 算法消耗——简单的距离比较就能消耗大量运算资源
- 更新难度——模型越大更新包就越大
- 符合客户的实际环境，迁移学习,现场训练？
- 神经网络的阈值调整反向传递？
———2U设备？4U设备？机柜整柜交付？

想得清楚：攻击链识别，助力事件还原



自动关联攻击链：
从外网用什么技术攻击了哪里，
然后分别攻击到了哪里

依据上下文得出结论

时间范围	攻击源	攻击目标	攻击类型	攻击次数	攻击成功率	攻击结果
2018-12-21 12:00:01 至 2019-01-21 12:00:01	Trojan[Exploit]/MSWord	111.001.132	木马	1	100%	成功
2018-12-27 15:54:19 至 2019-01-21 12:00:01	Worm[Quake]/Mykroware	111.001.132	蠕虫	3	100%	成功
2018-10-19 18:50:37 至 2019-01-21 12:00:01	Trojan[Exploit]/MSWord	111.001.132	木马	5	100%	成功

锁定攻击组织
描述攻击资源及装备
还原攻击过程

攻击阶段	攻击源	攻击目标	攻击类型	攻击次数	攻击成功率	攻击结果
攻击源	111.001.132	111.001.132	木马	1	100%	成功
攻击源	111.001.132	111.001.132	蠕虫	3	100%	成功
攻击源	111.001.132	111.001.132	木马	5	100%	成功

"Trojan[Exploit]/MSWord" 1个目的进行了传播 告警分析报告

核心行为: Exploit

攻击资源: MSWord

攻击过程: MSWord

攻击结果: MSWord



想得清楚：ATT&CK / NSACSS两大威胁框架，呈现攻击全貌



- 揭示威胁所处的攻击阶段；
- 呈现攻击方法和手段；
- 帮助建立对攻击链的完整认知；
- 帮助发现最新攻击技术，检测；
- 指导开展安全分析、响应，加强

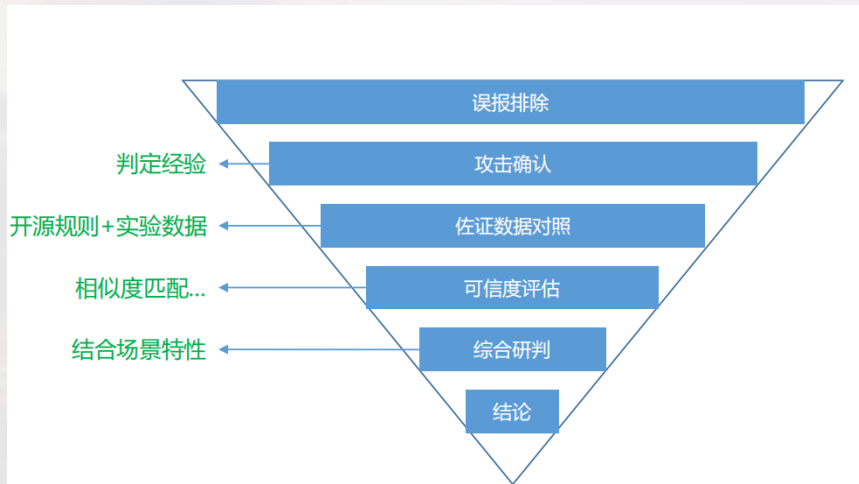
ATT&CK模型将攻击行为按照目的分为12个类别，并详细描述了各类别中具体的攻击技术。

威胁阶段

初次访问	执行	持久化	提权	防御规避	凭证访问	发现	横向移动	收集	命令控制	渗透	影响
	Graphical User Interface 1	Shortcut Modification 1		Obfuscated Files or Information 1		System Information Discovery 1				Data Encrypted 1 1	Runtime Data Manipulation
		Registry Run Keys/Startup Folder 1		Process Hollowing 1		Security Software Discovery 1					
				Process Injection 1 2		Process Discovery 1					
				Virtualization/Sandbox Evasion 1 1		Application Window Discovery 3					
				Position data 1		Peripheral Device Discovery 1					

借助威胁框架，告别单点式的载荷检测分析，统观全局，开展系统性的关联分析

想得清楚：威胁框架研判和攻击链识别的算力代价



风险资产识别的处理模型

- 基于中观理解上下文
 - 微观：单一告警事件
 - 宏观：网内整体情况
- 过去发生了什么，现在正在发生什么，未来将要发生什么
 - 过去：数据识别、索引、提取、老化
 - 未来：
 - 威胁情报匹配，TTPs支撑
 - 资产数据支撑
- 消耗：磁盘IO、内存缓存

说的明白：猎杀分析闭环，支撑APT组织追踪，绘制事件全貌



1、基于恶意代码精确分类的 C&C静态、动态提取方法分类为木马家族后的恶意代码，可基于还原出的文件实体，对对应家族进行静态快速的C&C解密，发现C&C地址。对不能静态解密的家族，可以基于沙箱进行动态分析，获取其C&C地址，同时，沙箱产生的pcap可用于生产流量检测、识别规则。

2、基于流量规则的C&C服务器识别方法基于产生的恶意代码通讯特征，可进行流量的控制端、被控主机识别。发现的C&C控制端，可针对性进行升级行为采集，从流量中捕获新的恶意代码。



说的明白：有效支持归纳攻击组织、资源及装备



恶意代码传播关联

多个IP地址传播了同一个恶意代码，发现同一攻击组织。

PAYLOAD相似性关联

根据攻击的payload的相似程度，发现属于同一个商业军火。攻击的payload被PTD发现后，留取pcap，可发送至平台。

恶意代码传输协议关联

基于同一恶意代码家族的放马协议的放马文件协议 (http、tftp、ftp)、路径、用户名、文件名进行关联。在PTD发现同一恶意代码家族的传输时，可同时提供对应协议的这些信息，供进行分析关联。此项目由PTD的全要素进行支持，同时由恶意代码的精确命名进行支持。

恶意代码C&C关联

发现多个恶意代码连接了同一个C&C服务器，发现同一攻击组织使用了这个恶意代码。

PAYLOAD放马源关联

根据攻击的payload，提取放马源的放马协议 (http、tftp、ftp) 进行关联，根据放马协议的路径、用户名、文件名进行关联。发现多个放马源及攻击源属于同一个攻击组织。

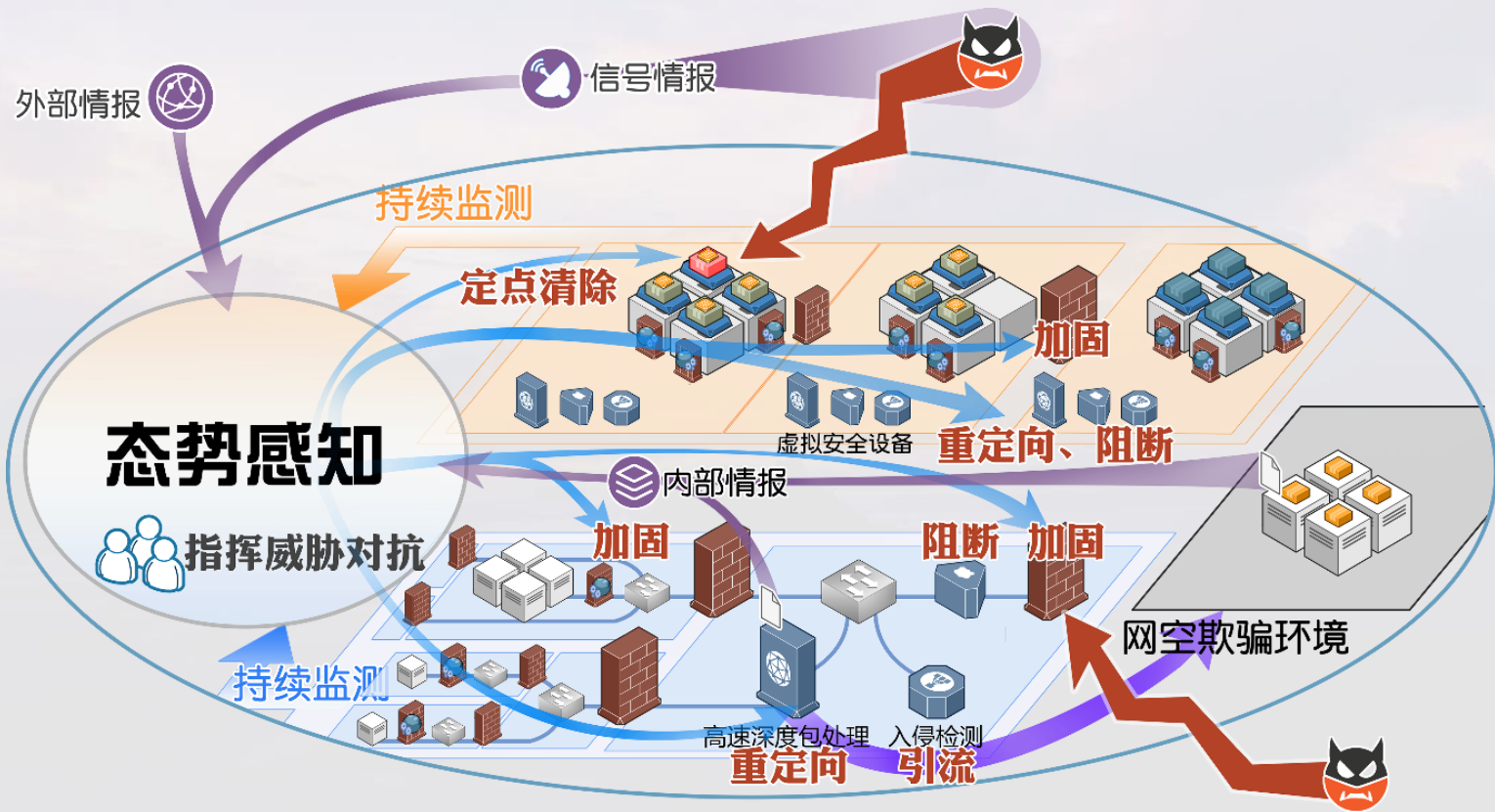
恶意代码通讯家族关联

基于检出恶意代码通讯家族、检出恶意代码C&C服务器的国籍及端口信息，发现攻击组织。

说的明白：威胁猎杀，分析扩线的算力代价

- 威胁猎杀的核心：建立假设——证实、证伪
 - 证据的加入——扩展问题边界，或缩小范畴
 - 响应时间——适合人类处理
 - 协同工作——多个分析会话，会话间协同。
 - 报告——数值、统计、列表
- 工作：
 - 原始数据：解析、分词、索引、检索、抽样、统计、分布.....
 - 分析过程：快速判断有无、抽取、步骤快照记录、撤销操作.....
- 算力需求
 - 技术：哈希、位图、红黑树、倒排索引、列式存储、压缩、LSM.....
 - 对CPU、磁盘存储、内存、IO的全面需求

可处置，策略执行



03

新变化—— 基于国产化载体的流量检测

预防“卡脖子”，推进国产改进势在必行

数字化需求

随着国内企业数字化进程推进，不断提出对适合中国市场的基础软件、应用软件、基础设施、信息安全的需求，催生出巨大的信创市场

产业升级

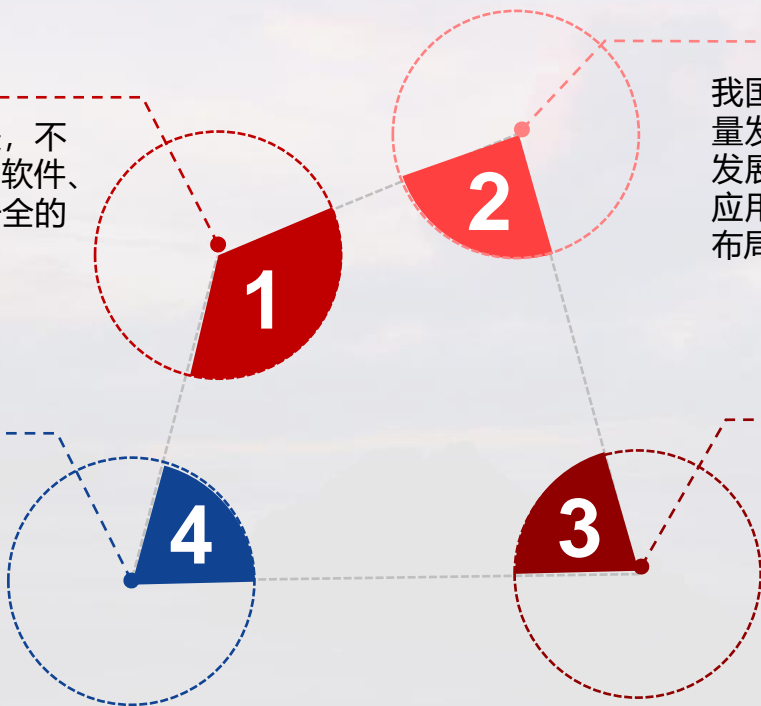
我国的经济由高速发展进入了高质量发展的新常态，为了推动高质量发展，我们就必须加快在基础软件、应用软件、基础硬件等创新领域的布局

信息安全

过去很多年间，我国IT底层标准、架构、生态等多由美国公司制定，因此存在着诸多安全风险。为了保护我国信息安全，建立自主可控的信息网络，国家提出了信创的要求

国际环境变化

关键核心技术是国之重器，在大国博弈的背景下，摆脱受制于人的局面，必须解决关键核心技术领域的技术缺失，坚定走自主创新之路。



追随自主创新号角，全面兼容，提升安全可控能力



安天积极支持自主创新战略，助力推进国产产业、网络安全领域的深度融合。

探海在流量检测响应领域已推出基于龙芯、兆芯、飞腾、海光自主可控平台的设备。



国产化性能及功耗有长足进步，但算力缺口依然存在

国产CPU性能盘点

公司	CPU	主频	核心数	单核性能	工艺	时间	备注
申威	SW3232	2.4G	32	25+	14/16nm	2020	32核rate 600+
	SW432	2.4G	4	25+	14/16nm	2021	
龙芯	3A5000	2.5G	4	25+	12nm	2020	单核25至30
	3C5000	2.5G	16	25+	12nm	2021	
飞腾	FT-2000/4	2.6G	4	16.5	16nm	2019	
	D2000	2.3G	4	14.6	14nm	2020	
	新一代服务器CPU	?	64至128	25+	?	?	64核rate 1000+
兆芯	KX-6000	2.7G	4、8	18+	16nm	2019	
	KH-40000	3G	32	25+	16nm	2021	32核rate 1000+ (ICC)
	KX-7000	3G	?	25+	16nm	2021	
海光		3.5G	8、32	35+	14nm	2018	

数据来源：
Eefocus测试数据

单核性能测算方法：
软件：SPEC CPU 2006

SPEC组织推出的CPU系统性能评估
测算

国产CPU典型功耗：100W-150W
(飞腾S2500、申威SW1621等)

典型Intel的CPU单核性能：40-50

机器学习需求的浮点数计算——一个不恰当的对比

机器学习的算力消耗：浮点矩阵计算

网络安全的算力消耗在哪？

流追踪、解码

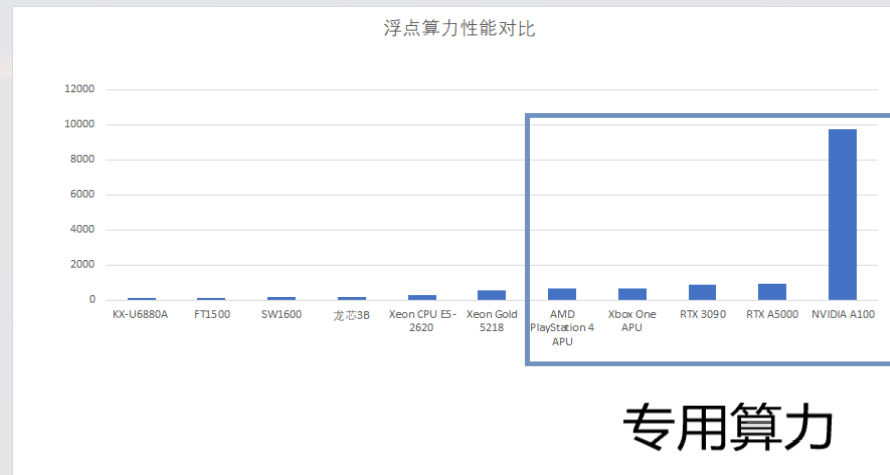
哈希、浮动关键字、正则表达式？

模型抽取、匹配？

高CPU、高内存，SIMD应用困难。

厂家	型号	单精度浮点数性能	双精度浮点数性能
兆芯	KX-U6880A		192 GFLOPs.
飞腾	FT1500		144 GFLOPS
申威	SW1600		140.8 GFLOPS
龙芯	3B系列	256 GFlops	128GFlops
Intel	Xeon CPU E5-2620		268.8GFlops
Intel	Xeon Gold 5218		665GFlops
AMD	AMD PlayStation 4 APU	1.84TFlops	
微软	Xbox One APU	1.31 Tflops	
Nvidia	RTX 3090	35.58 TFLOPS	556.0GFLOPS
Nvidia	NVIDIA RTX A5000	27.77 TFLOPS	867.8 GFLOPS
Nvidia	NVIDIA A100	19.49 TFLOPS	9.746TFLOPS

数据来源：新闻稿及公开数据收集





网络空间威胁对抗与防御技术研讨会
暨 第九届安天网络安全冬训营

安天 | 智者安天下

04 资源代价的一些思考

他山之石和不便之处



通用处理器

核心技术：NUMA、SIMD、网卡offload等

处理能力增长：更多核心+SIMD指令

价格降低：摩尔定律

限制：显著慢于流量增长速度



SmartNIC

核心技术：片上FPGA、高速片上互联、并行计算

处理能力增长：高效片上算法

价格降低：流量NIC领域发展

限制：IP核、时序处理、片上内存、并行计算
匹配表和安全所需量级差距太远
发展方向偏向SSL Offload，而非匹配



GPU

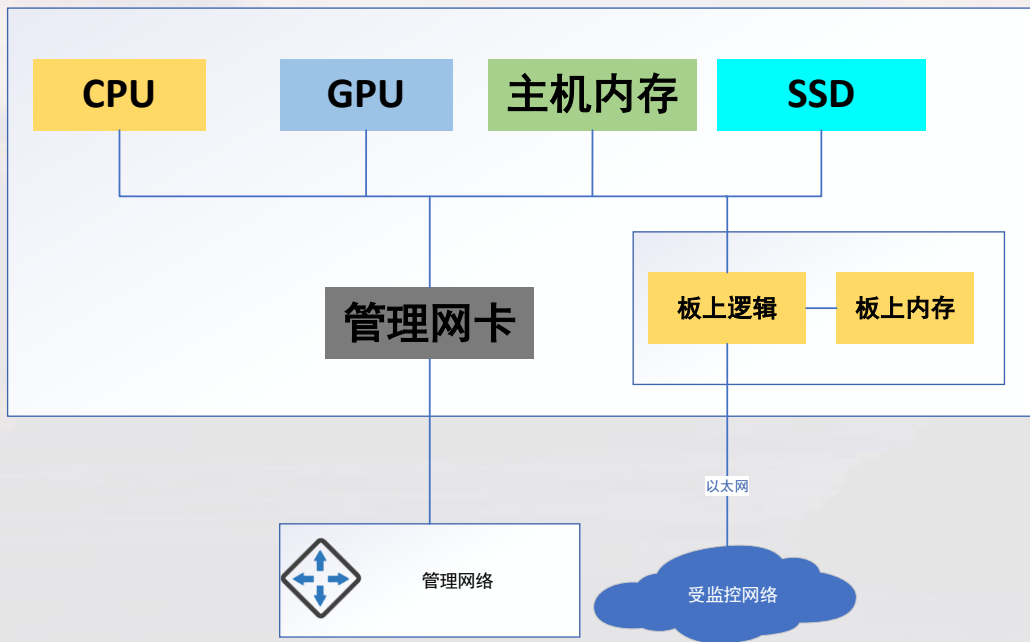
核心技术：并行计算、内存RDMA

处理能力增长：并行度、GPU核心数

价格降低：GPU制造工艺

限制：GPU片上内存太小
数据包拷贝、总线带宽

未来，我们一直在努力！



- 安全算法
- 特征匹配引擎
- 浮动关键字
- 情报匹配算法
- 流offload
- 加密解密
- 压缩解压
- 正则表达式

CPU、内存、主板为国产化的环境下，在单一设备上达成100Gbps骨干网流量处理，具备一定的复杂规则处理能力的设备



网络空间威胁对抗与防御技术研讨会
暨 第九届安天网络安全冬训营

亂雲飛渡

谢谢大家



安天冬训营 wtc.antiy.cn