大数据安全产品实践随谈

alan.qian@huawei.com

OWASP 中国

The Open Web Application Security Project

About Me

OWASP 中国
The Open Web Application Security Project

威胁源于未知
安全源于感知

- 企业BG交换机与企业通信产品线
- 首席安全架构师兼安全防御TMG主任
- 负责安全业务领域的技术规划、安全产品核心引擎架构与安全智能中心构建，以及前沿安全技术研究
- 关注安全智能技术与安全生态建设
- 近期研究兴趣是开放威胁情报共享机制

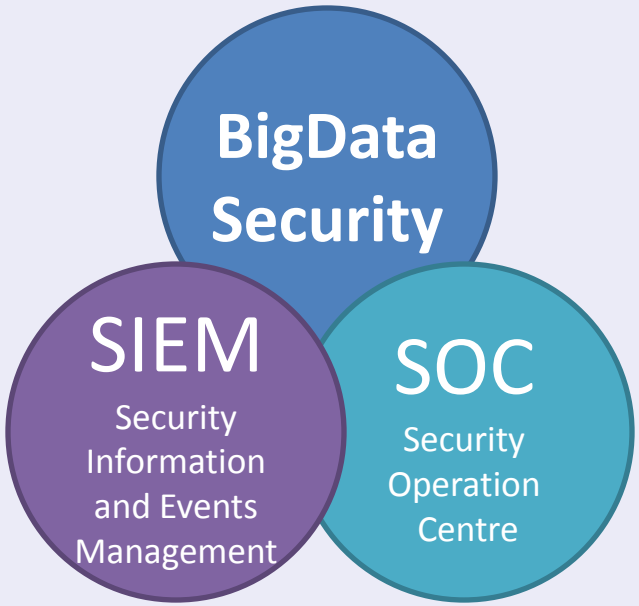
Agenda

OWASP 中国
The Open Web Application Security Project

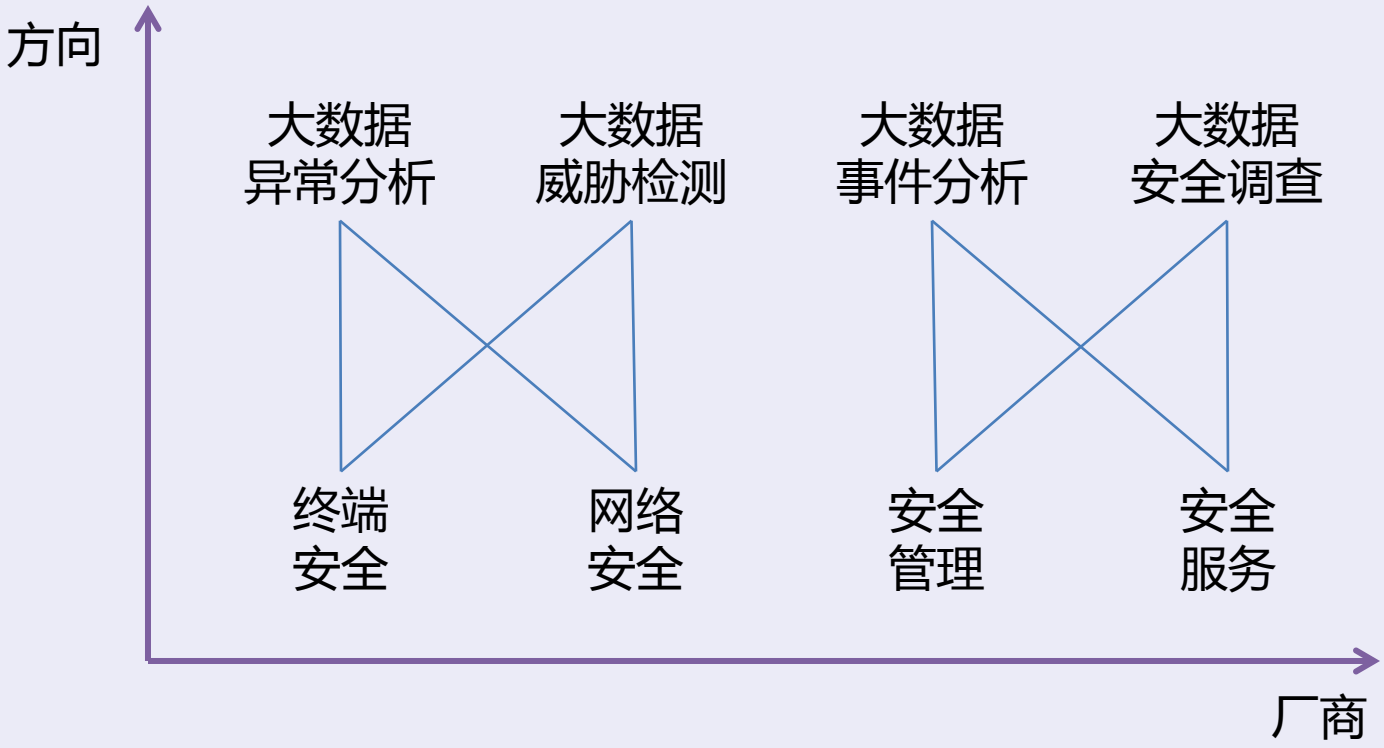
- 定位
- 技术路线
- 技术要求
- 数据要求
- 处理流程
- 关键技术
- 产品部署
- 未来挑战

定位

- 日志
- 样本
- 威胁情报
- 环境数据



技术路线



APT对抗的策略要求

已知威胁	未知威胁
有明确的特征	新型行为特点
可快速提取行为特征	行为特征隐秘复杂，不易定义
在非法场景内出现	在合法场景内发生
表现出非授权使用	表现出授权使用
常发生在基线之外	常发生在基线之内
在监控能力之内	在监控能力之外
样本精确分析	大数据持续分析

技术要求

OWASP 中国
The Open Web Application Security Project

安全分析

流量学习
DPI应用识别
IPS/IDS入侵检测
AV/沙箱分析
访问控制/认证审计
内容过滤
安全信誉
威胁情报智能

数据科学

大数据技术
数据规划与管理
数据可视化
统计分析
关联分析
机器学习
算法模型

攻防思维

学习型组织
黑客思维
创新能力
分析能力
红蓝对抗演练

技术要求

OWASP 中国
The Open Web Application Security Project

人工分析
喜欢模式，设定规则，善于启发，创造性好，发现银弹，易出错，连续性不好

机器分析
可训练的，复杂度高，大计算量，容易扩展，高精度，连续性好

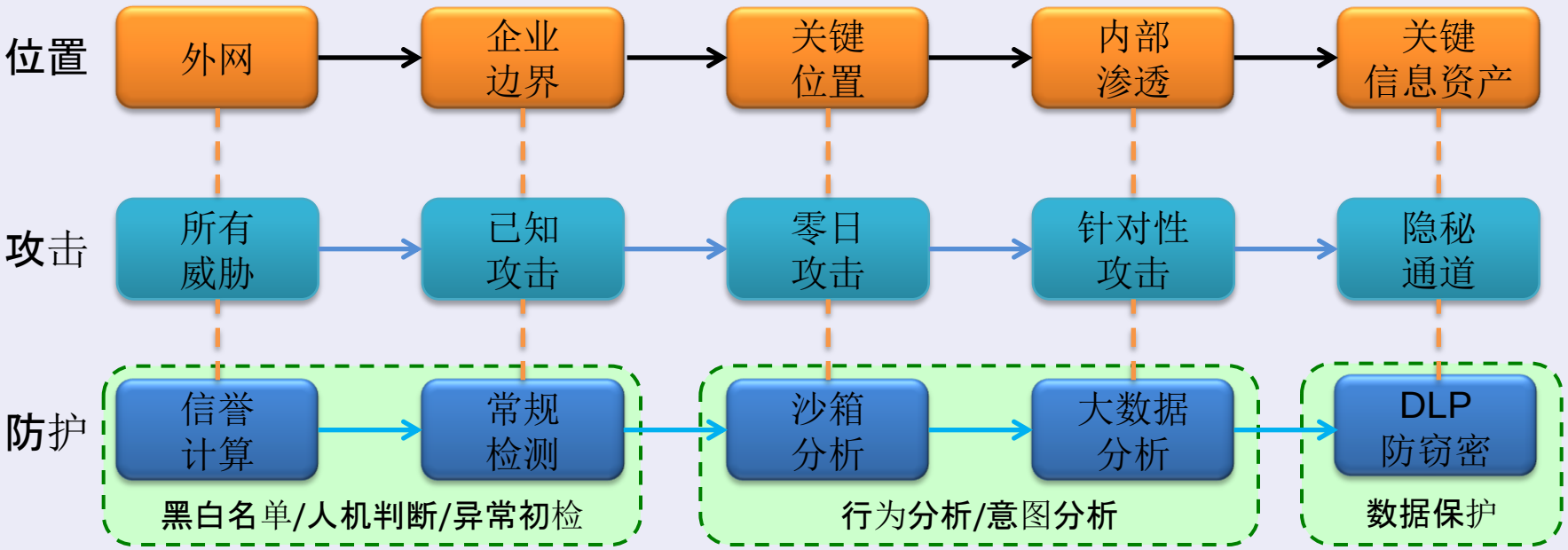
主机数据
AntiVirus
HIDS/HFW
DLP
主机AAA
终端探针

网络数据
Firewall/IPS/IDS
AntiSpam/AntiPhishing
Content Filtering
沙箱分析
流量审计
流探针

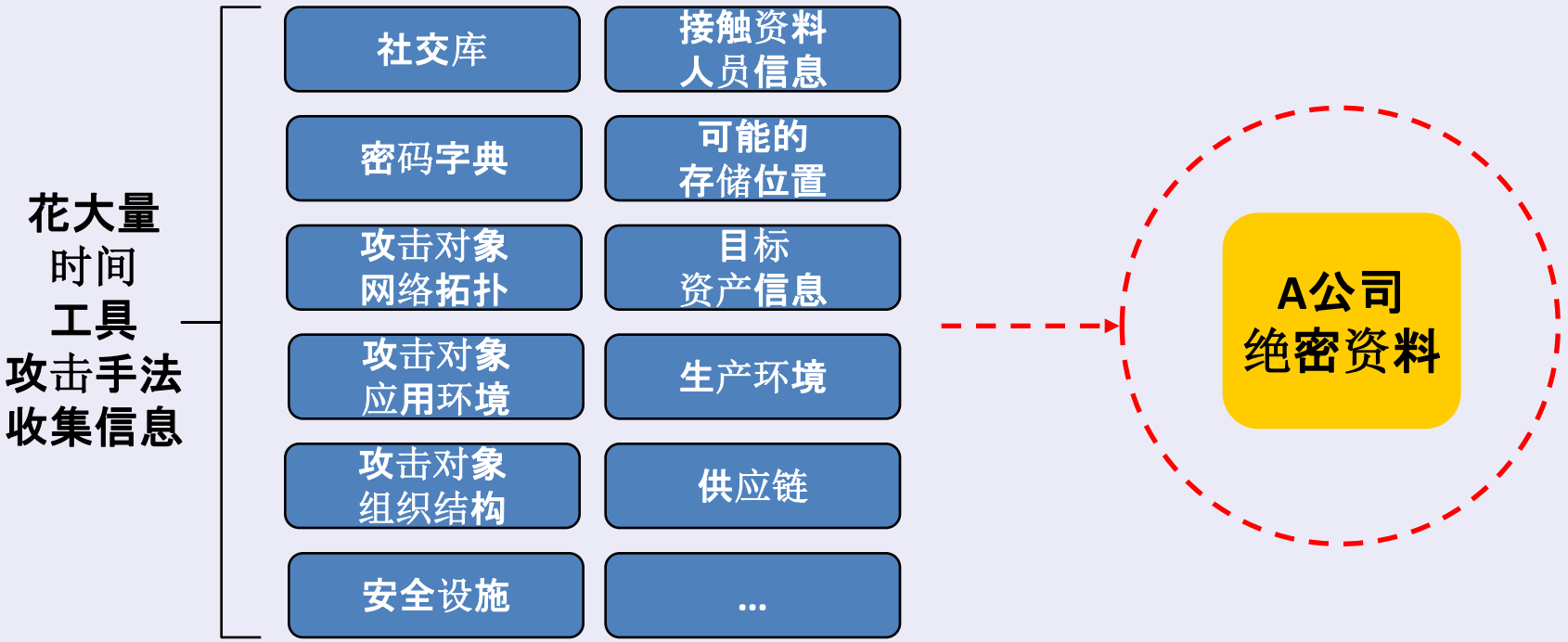
环境数据
弱点扫描器
网管/控制器
网络爬虫/虚拟蜜网
人工采集
漏洞库
社区反馈

技术要求

OWASP 中国
The Open Web Application Security Project

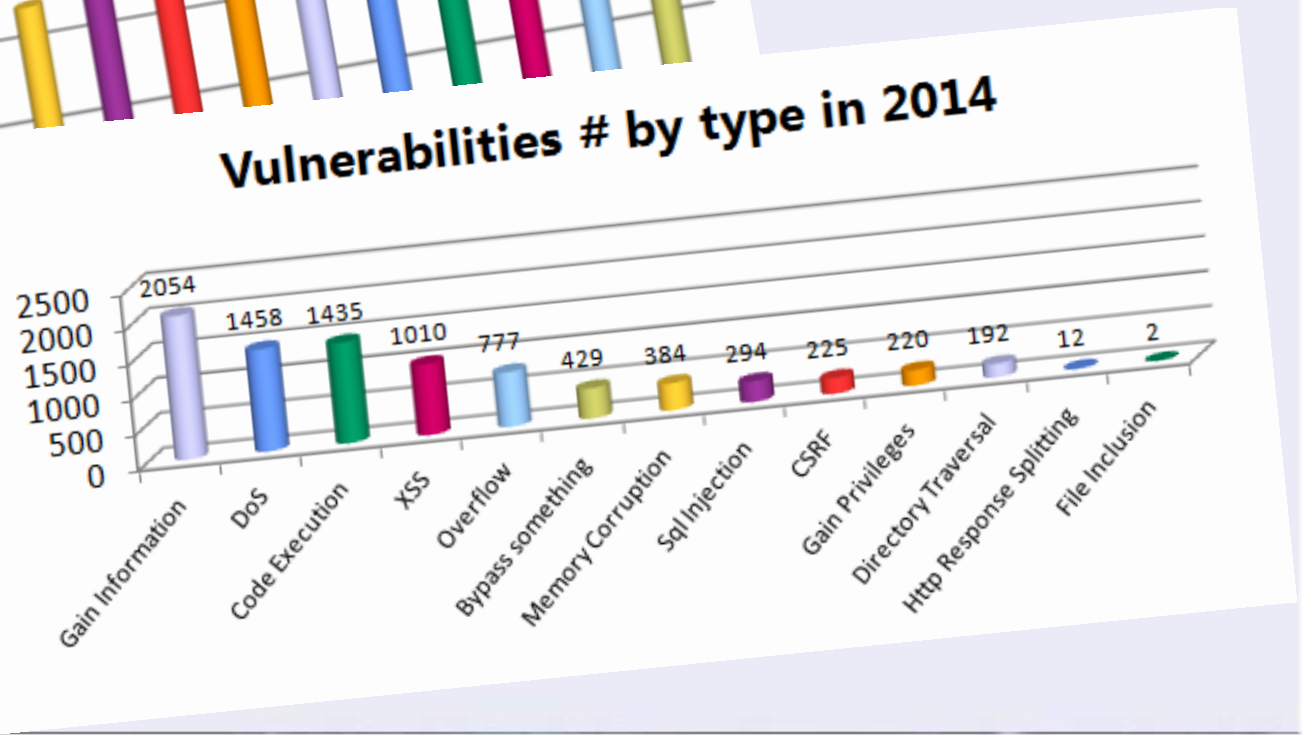


数据要求


花大量
时间
工具
攻击手法
收集信息
社交库
接触资料
人员信息
密码字典
可能的
存储位置
攻击对象
网络拓扑
目标
资产信息
攻击对象
应用环境
生产环境
攻击对象
组织结构
供应链
安全设施
...
A公司
绝密资料

数据要求

OWASP 中国
The Open Web Application Security Project



数据要求

OWASP 中国
The Open Web Application Security Project

背景数据	**前提、假设、条件等**
来源数据	攻击者、误用者、故障源
对象数据	被攻击者/攻击目标/破坏对象
环境数据	攻防所处的计算环境、网络环境、物理环境、地理环境
内因数据	脆弱性/安全漏洞/配置错误
模式和方法	伪装特征、代码特征、躲避特征、主机行为、网络行为、传播模式
过程数据	攻击路径时序：人/网络/主机
结果数据	威胁的可能性 威胁的程度与范围

数据要求

OWASP 中国
The Open Web Application Security Project

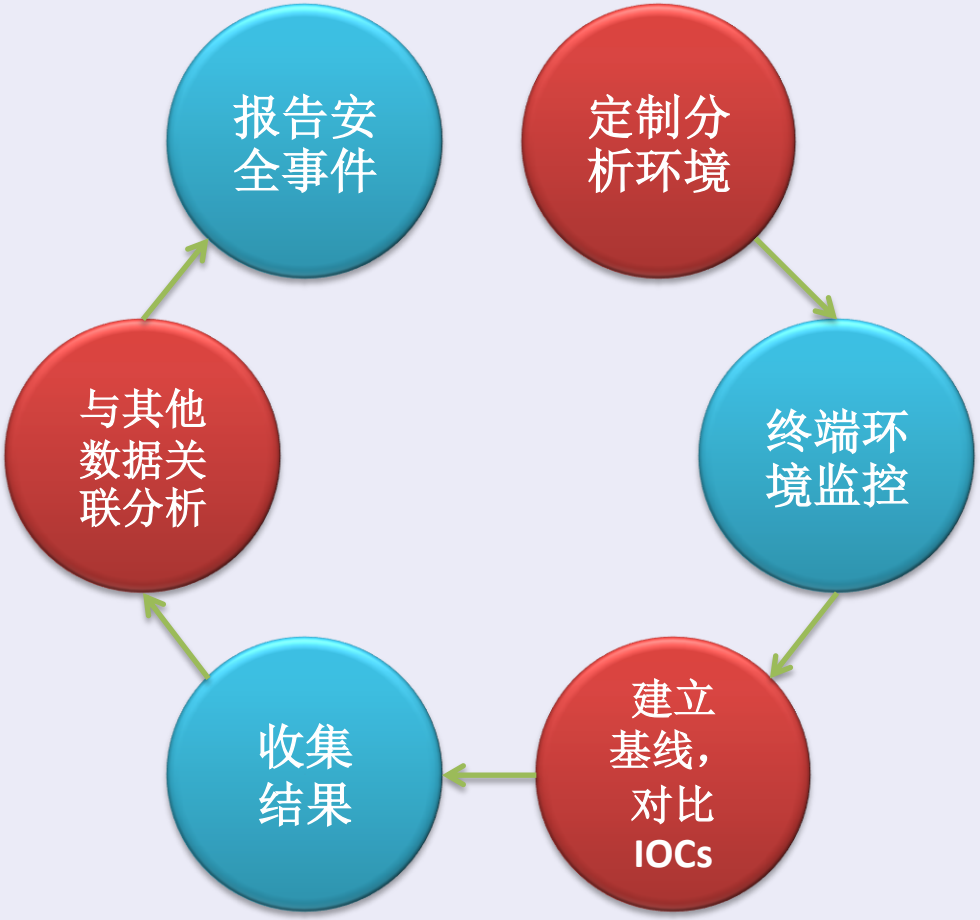
数据要求

OWASP 中国
The Open Web Application Security Project

威胁情报

——

权威组织发布
数字安全社区
独立研究者
开放论坛
商业合作



沙箱分析

——

进程分析
注册表监控
网络连接
DLL加载
环境变量
文件操作
软件配置
安全软件

用户要求

OWASP 中国
The Open Web Application Security Project

精准

- APT攻击链检测
- 可疑行为+基线+大数据+机器学习+综合评分

高效

- 攻击行为路径可视化
- 首次感染
- 秒级智能检索
- 情报

智能

- 按需采集、资源池
- 安全隔离、修复、限制

处理流程

OWASP 中国
The Open Web Application Security Project

采集

日志事件

终端行为

Netflow

全流量

检测

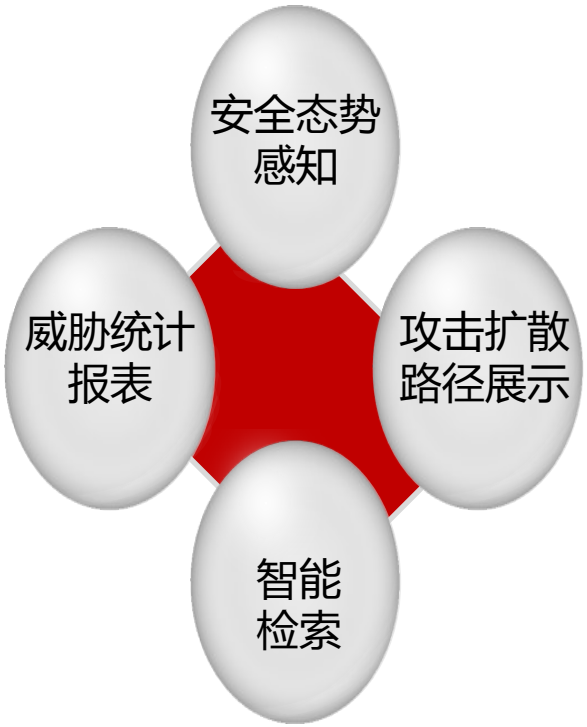
威胁分析与检测

签名过滤 | 行为检测 | 基线学习 | 机器学习 | 多维关联

大数据建模分析

贯穿APT攻击全阶段检测

呈现



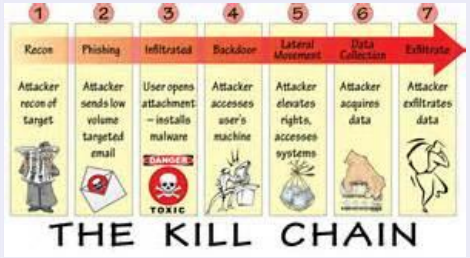
基于Hadoop 的大数据存储分析平台

处理流程

OWASP 中国
The Open Web Application Security Project

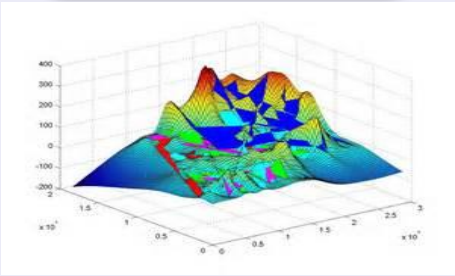
全攻击链持续分析

采集全网数据
汇总检测结果
多维分析对照
调用云端分析能力
攻击链跟踪
围绕核心资产做全量分析



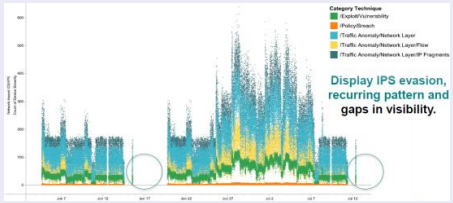
可视化威胁建模

构建可视化建模工具，根据分析任务，进行数据预处理，基于更小的数据集与数据维度进行分析算法设计



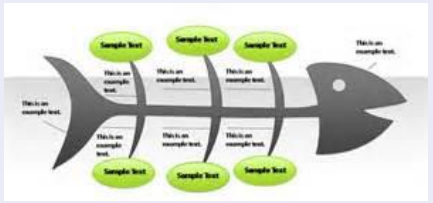
威胁可视化

利用机器学习与可视化技术，将数据中隐藏的异常状态呈现出来，便于机器与人工进行威胁挖掘与风险判定



大数据安全调查

基于大数据智能搜索技术，针对上报的异常或威胁事件，在海量数据中进行高性能安全调查，实现快速判定与响应



威胁场景

OWASP 中国
The Open Web Application Security Project

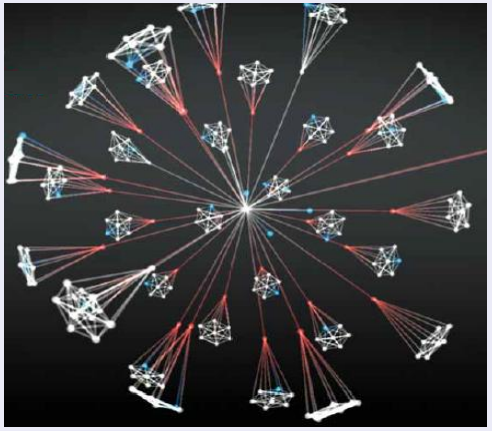
	初步感染	创建驻点	提升权限	内部侦查	完成任务
常见手段	•社会工程 •定制恶意软件 •鱼叉式攻击 •水坑式攻击	•下载恶意软件 •多个CC通讯 •第三方应用漏洞利用	•盗取凭据 •密码破解 •绕过校验	•侦查关键系统 •系统&活动目录账号枚举	•整合数据 •盗取数据 •日志擦除
异常行为	•异常扫描行为 •异常文件行为 •可疑钓鱼邮件	•异常动态域名连接 •周期性心跳连接 •未知恶意行为	•SSH 破解 •高级账号撞库破解	•可疑扫描 •可疑服务（协议+端口） •可疑Http/Https服务	•隐蔽通道行为 •加密通道行为 •异常流量行为

威胁可视化

原始日志

情报

优先级



用户感知

应用感知

地域感知

关键资产日志

情报+优先级

攻击扩散路径

丰富上下文信息

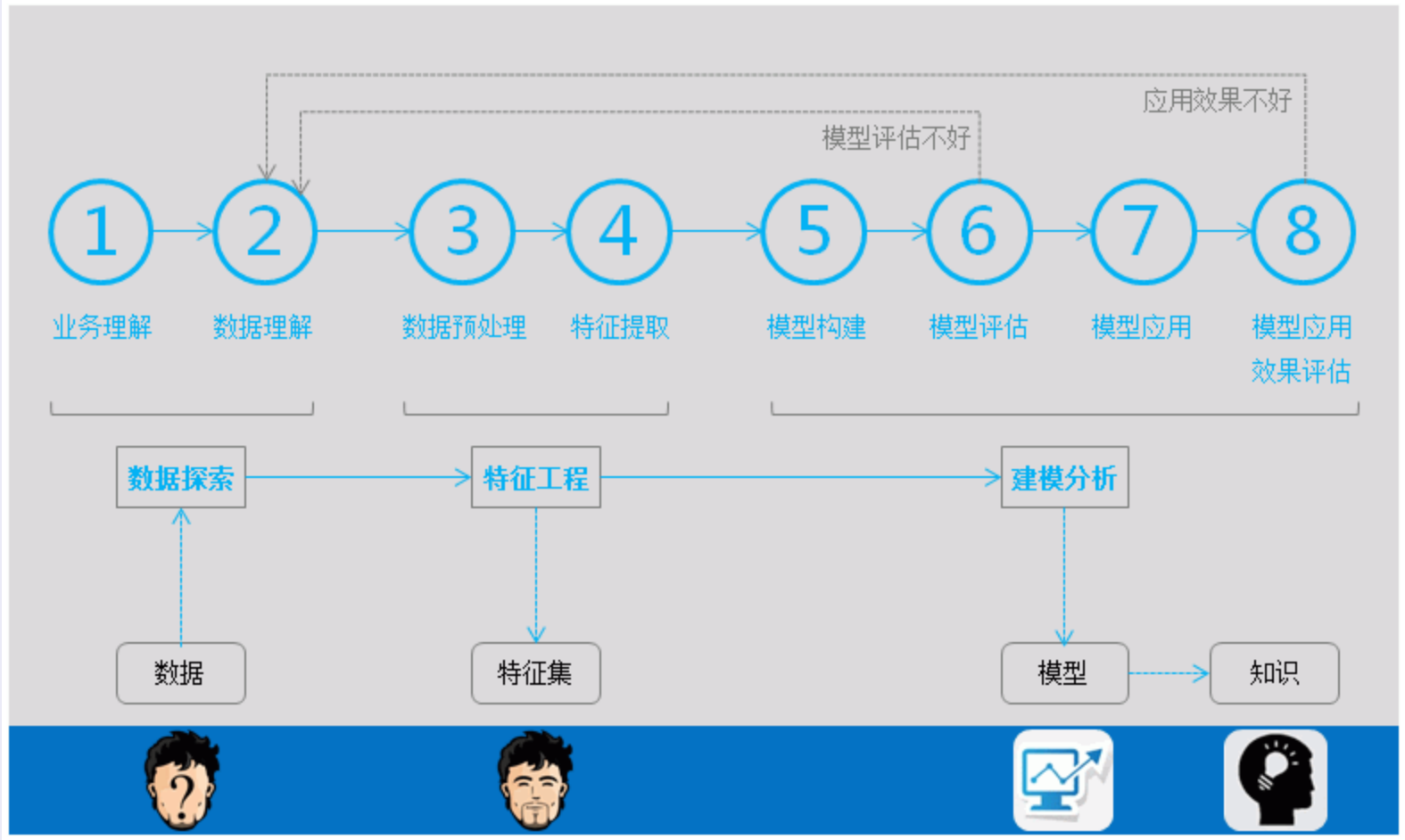
可视化威胁建模

OWASP 中国
The Open Web Application Security Project

FusionInsight Miner

海量数据 繁杂多样	图形化建模 | 多人协作 | 操作简单 分布式存储 | 并行计算	有价值的知识 产品特征

可视化威胁建模


应用效果不好
模型评估不好
1
2
3
4
5
6
7
8
业务理解
数据理解
数据预处理
特征提取
模型构建
模型评估
模型应用
模型应用
效果评估
数据探索
特征工程
建模分析
数据
特征集
模型
知识

可视化威胁建模

OWASP 中国
The Open Web Application Security Project

步骤	挖掘阶段	各阶段做什么	Miner如何实现	输入/输出
1	业务理解	制定目标和计划，希望从数据挖掘中得到什么。	数据探索	数据
2	数据理解	收集与分析任务相关的数据，增加对数据的说明，并关注数据的质量。		
3	数据预处理	将数据变换或统一成适合挖掘的形式。	特性工程	特征集
4	特征提取	抽象公共部分，形成特征工程。		
5	模型构建	通过拖拽及可视化调整完成建模过程。	建模分析	模型、知识
6	模型评估	根据某种兴趣度度量，识别提供知识的真正有趣的模式。		
7	模型应用	将应用数据输入模型，输出结果，从中获取有用的知识。		
8	模型应用效果评估	将模型输出的结果与现实情况进行对比。		

可视化威胁建模

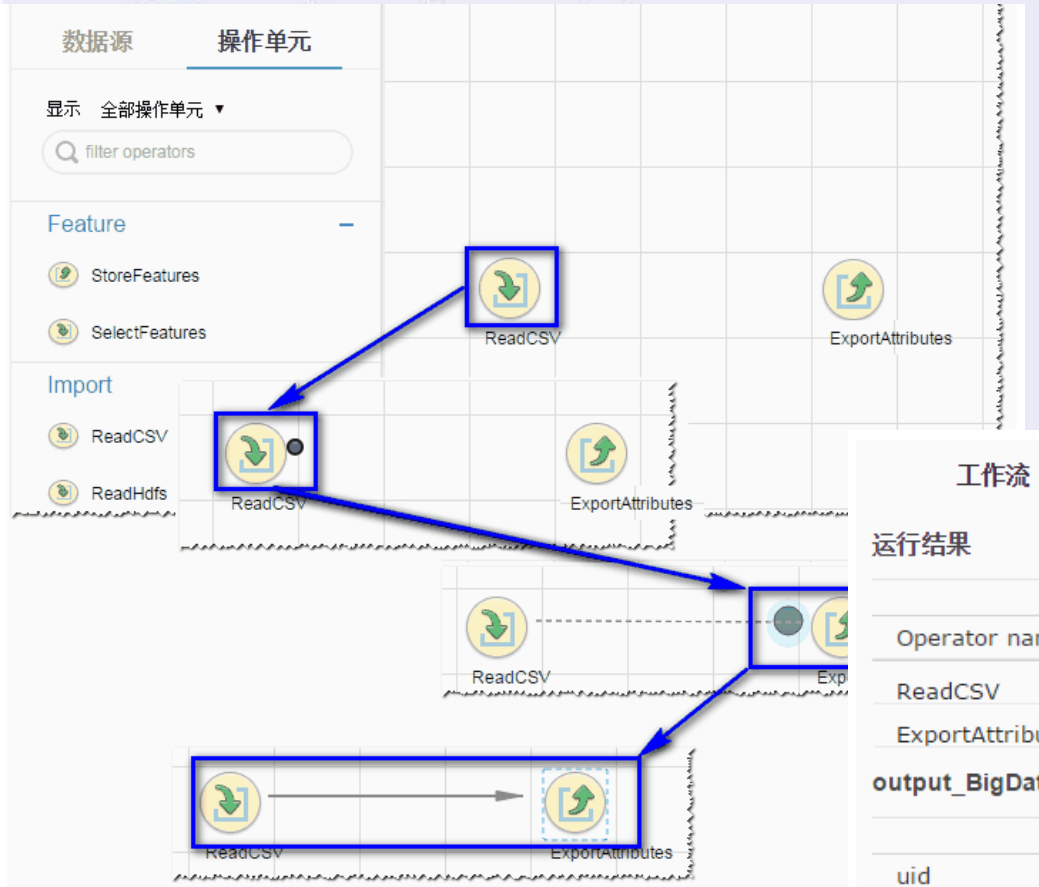
OWASP 中国
The Open Web Application Security Project

在认知心理学领域中，人在解决问题时要利用各种算子来改变问题的起始状态，经过各种中间状态，逐步达到目标状态，从而解决问题。解决问题中的种种操作被称为算子（Operator）。
在Miner建模分析时，算子代表某个分析子步骤，屏蔽编程细节，直接在工作流画布中拖拽算子、连接算子和修改算子属性，实现对数据的导入、导出、转换等处理。

算子类型	算子名称	算子用途
Feature	• StoreFeatures • SelectFeatures	• StoreFeatures 生成特征集数据 • SelectFeatures 选取特征数据
Import	• ReadCSV • ReadHdfs • ReadHive • ReadModel	• ReadCSV 从HDFS中导入CSV文件及其对应的描述文件 • ReadHdfs 从HDFS读取指定文件或者文件夹的内容 • ReadHive 从Hive中读取一个数据表 • ReadModel 从数据库中读出模型索引，并呈现给用户
Export	• ExportAttributes • PersistHdfs • PersistHive • PersistModel • PersistView • WriteCSV	• ExportAttributes 导出Hive表的属性到HDFS文件中 • PersistHdfs 将文件或文件夹数据写入到指定的HDFS路径中 • PersistHive 将输入的行列式数据写入到指定的Hive表中 • PersistModel 保存训练出来的模型 • PersistView 将其它算子输出的视图数据持久化 • WriteCSV

可视化威胁建模

OWASP 中国
The Open Web Application Security Project



工作流 运行结果 运行日志

运行结果

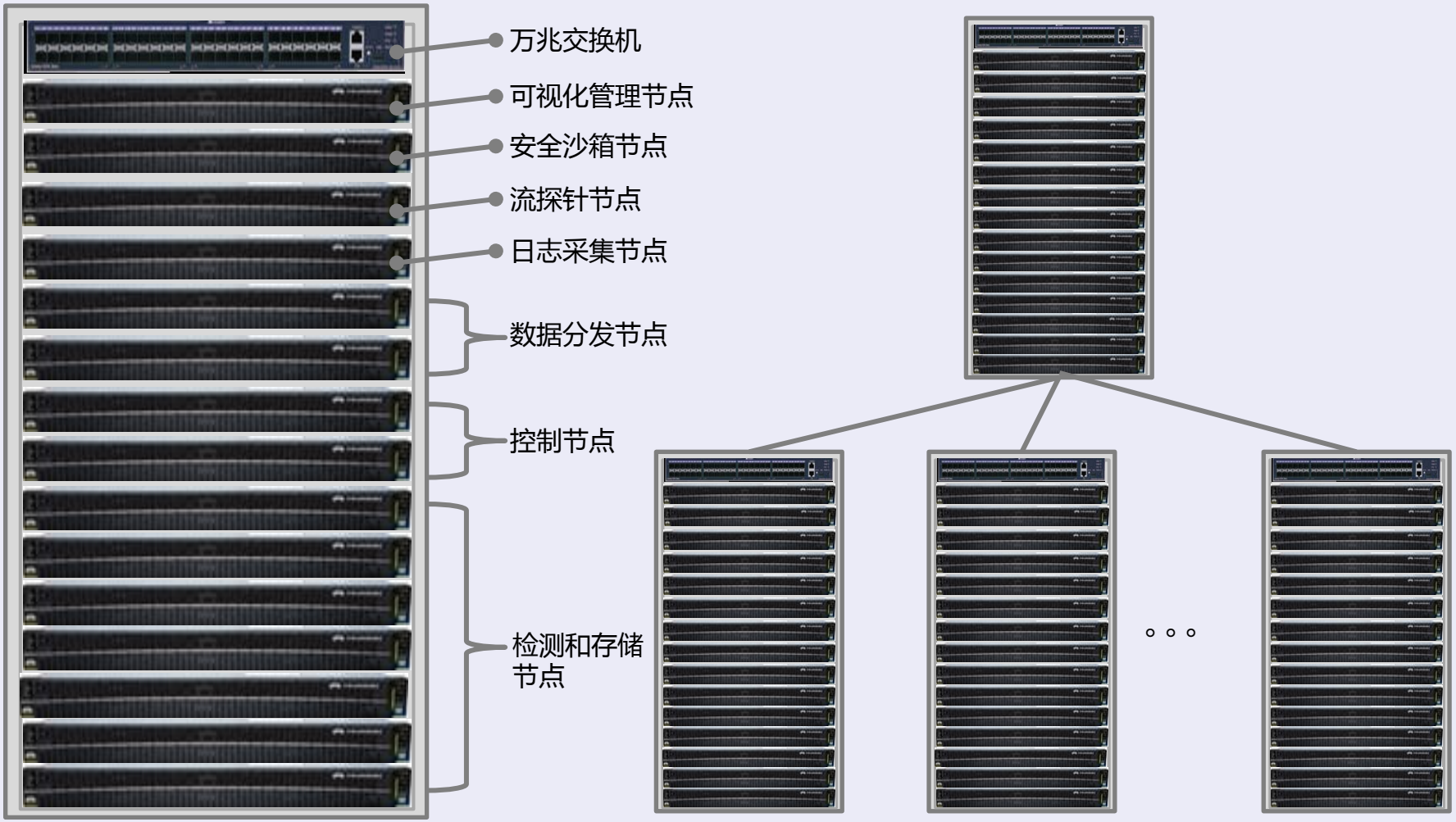
Operator name	Status	Results
ReadCSV	SUCCESS	output_BigDataSet
ExportAttributes	SUCCESS	

output_BigDataSet

uid	bigschool	single	phone
zha	xiada	NOT	181
liuz	shifan	NOT	188
hou	youzhen	YES	130
gac	zhongnan	NOT	186

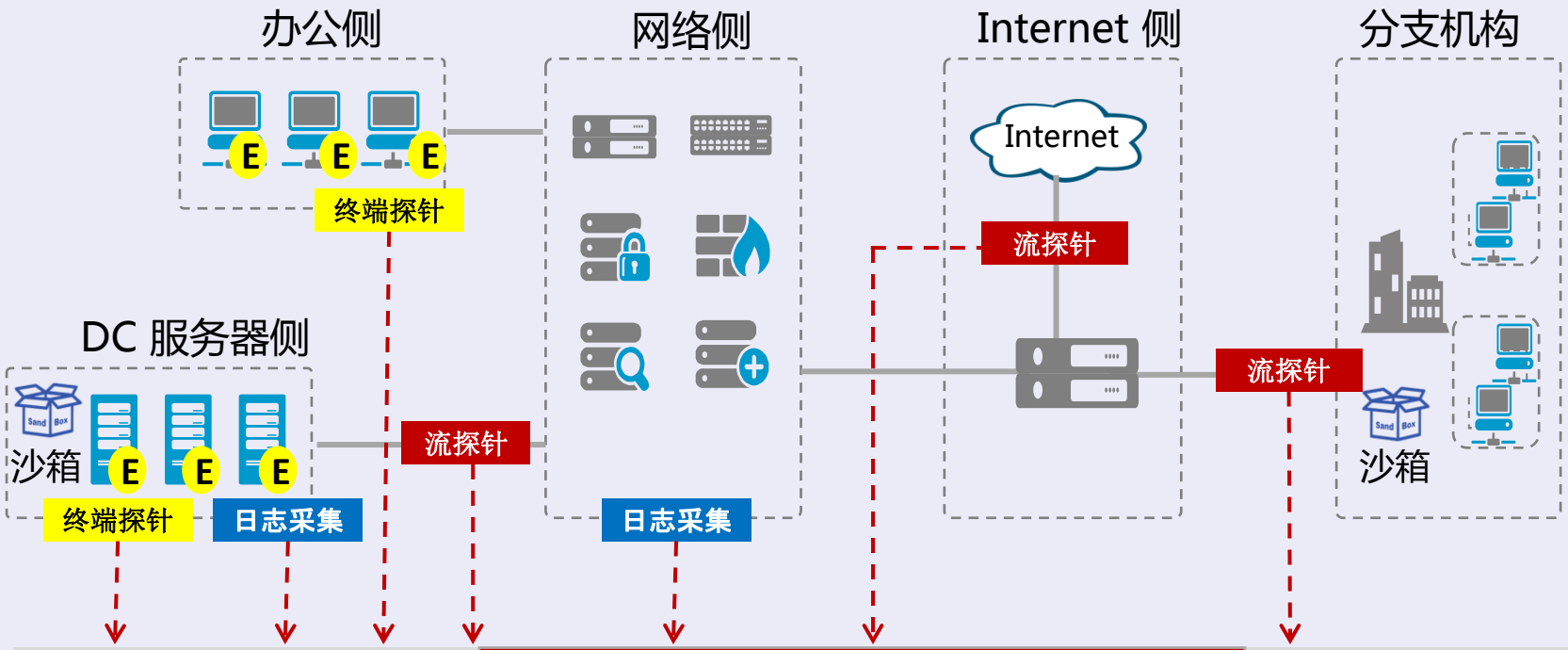
产品形态

OWASP 中国
The Open Web Application Security Project



产品部署

OWASP 中国
The Open Web Application Security Project



未来挑战

OWASP 中国
The Open Web Application Security Project

- 求异思维：寻求否定之否定
- 迷宫模式：入口_@$%&*出口
- 实用主义：“意有定向，招无定式”
- 反功能：misuse, abuse
- 隐藏与混淆
- 拟人拟态
- 社会工程学

编程大师说：“任何一个程序，无论它多么小，总存在着错误。”

初学者不相信大师的话，他问：“如果一个程序小得只执行一个简单的功能，那会怎样?”

“这样的一个程序没有意义，”大师说，“但如果这样的程序存在的话，操作系统最后将失效，产生一个错误。”

但初学者不满足，他问：“如果操作系统不失效，那么会怎样?”

“没有不失效的操作系统，”大师说，“但如果这样的操作系统存在的话，硬件最后将失效，产生一个错误。”

初学者仍不满足，再问：“如果硬件不失效，那么会怎样?”

大师长叹一声道：“没有不失效的硬件。但如果这样的硬件存在的话，用户就会想让那个程序做一件不同的事，这件事也是一个错误。”

没有错误的程序世间难求。

[Geoffrey James 1999 《编程之道》]

OWASP 中国
The Open Web Application Security Project

互联网安全+

Thanks