
ISC
2015
中国互联网安全大会
360互联网安全中心
ISC
2015
数据驱动安全
2015 中国互联网安全大会
China Internet Security Conference

分布式前置机器学习
在威胁情报中的应用

Application of Distributed Front-end Machine Learning in Threat Intelligence

cirrus gate

类似的尝试

Moving Big-Data Analysis
from a ‘Forensic Sport’ to a ‘Contact Sport’
Using Machine Learning and Thought Diversity

AJ Ferguson, NM Evans Harris
Information Assurance Directorate
National Security Agency
Journal of Information Warfare (2015) Volume 14

IAD is responsible for NSA's defensive mission and is widely acknowledged for leading innovative security solutions.

疑问：逆潮流？

大数据
分析平台

更清楚"看见"

分布式

前置

争分夺秒的时间竞赛

icloud-ios-appeid.us

iCloud账户钓鱼，巨大利润空间

注册：2015-09-26

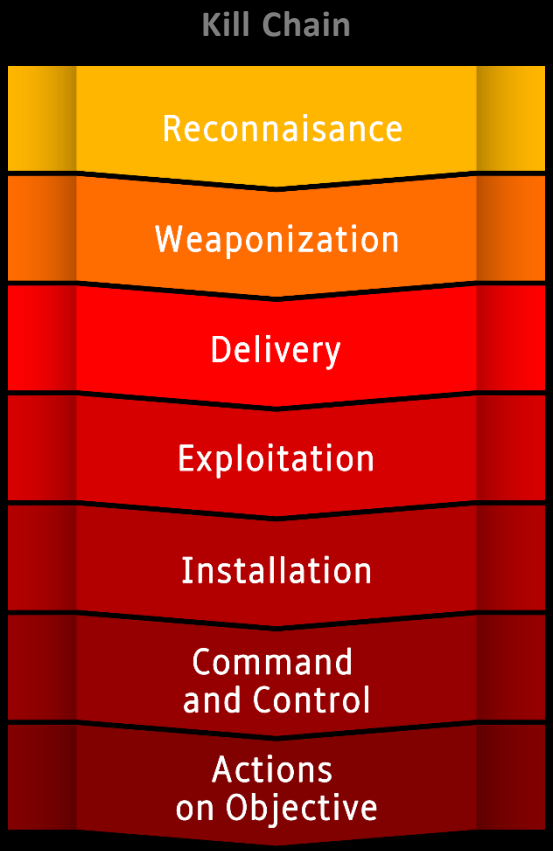
下线：2015-09-29

威胁情报有效期：3 天

机器识别：恶意域名+网页内容

定向攻击企业也可使用类似手法

尽可能在攻击链早期发现



- 争取防御空间和时间
- 对策1：利用威胁情报
- 对策2：在接触点实时分析
- 机器学习引擎前移
- 缓解大数据平台的滞后
- 缩短应急响应时间
- 不可能预警所有威胁

威胁情报生产 = 发现未知威胁

利用威胁情报 = 跳过未知威胁分析过程

自古华山一条路?

大数据

生产

威胁情报

正确道路但耗资巨大

基础设施不堪重负

Too much data?

永远塞车的带宽

日益拥挤的存储

耗时的查询分析

态势感知：借助威胁情报超越SIEM

MITRE

Situation Awareness

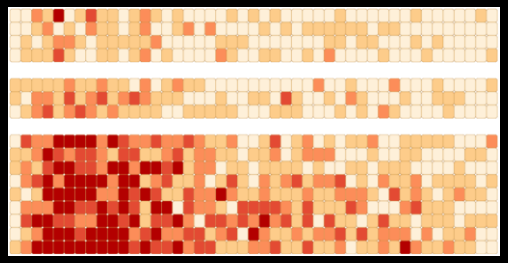
- It’s not a matter of simply aggregating all the tactical-level information available. Instead, status information must be correlated to the context of the mission or business, thus exposing the real impact to its operations.

- At the strategic level, it's important to be able to look well beyond simple incident data to identify threat actors, recognize trends in their activities, and expose their malicious objectives.

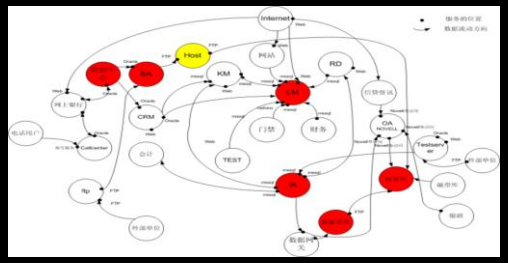
以价值为导向：业务、风险

- Too much data?
- Data is useless unless it drives decisions.

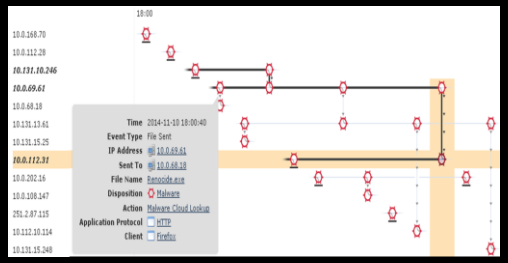
关键数据风险分布



业务系统行为异常



木马传染扩散路径



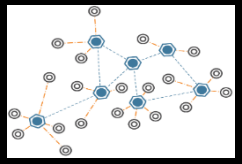
分布式前置引擎的能力

正确保留情境信息的情报最有价值：关联性、行动性、预测性

数据分类

木马分类

恶意域名

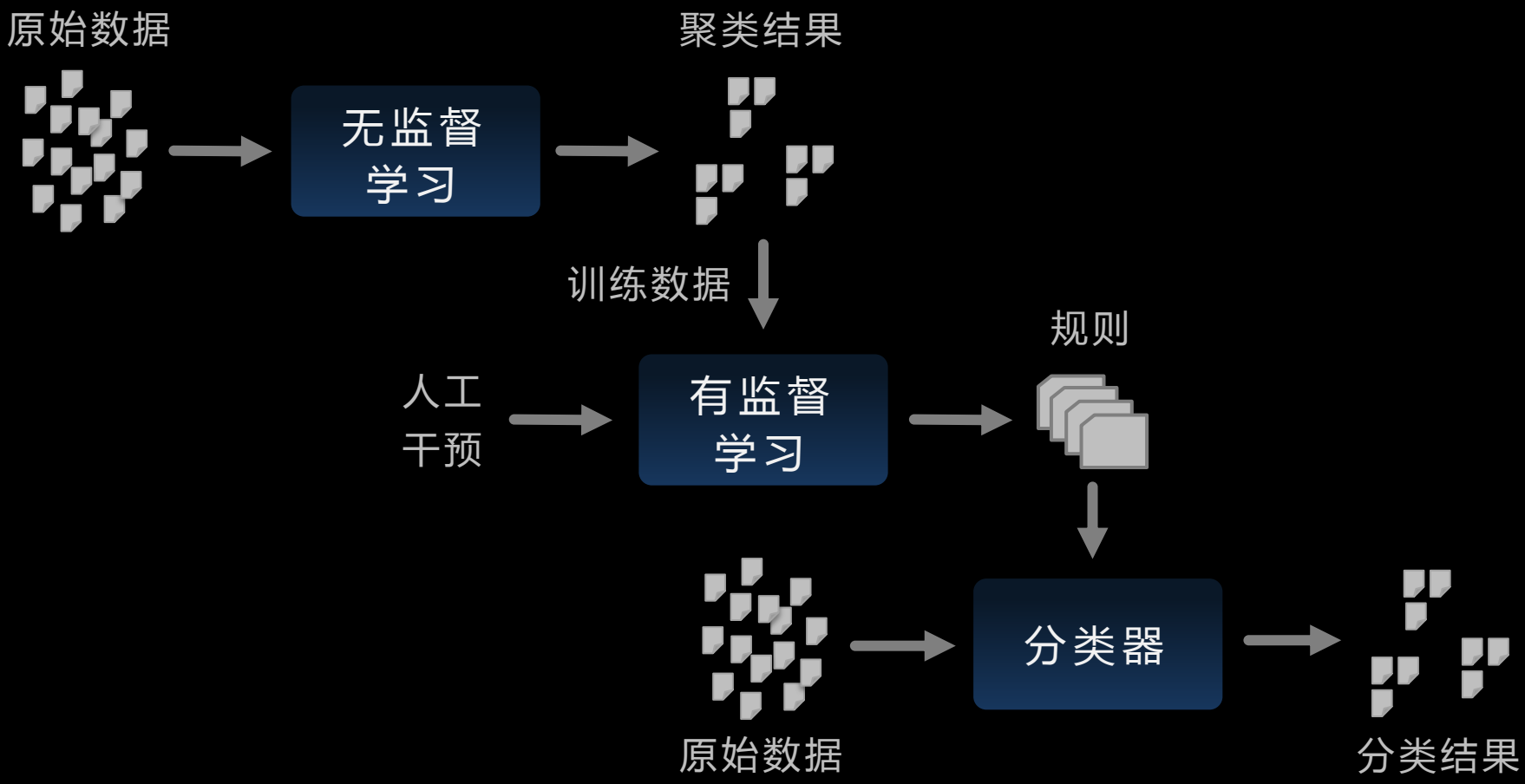


行为异常

网址分类

流量异常

机器学习



木马聚类分类

传统手工分析

- 功能列表
- 编程错误
- 加密方式
- 免杀对抗手段
- 进程设计
- 释放文件模块
- 通讯方式

机器学习选取特征

- 代码复用
- API调用
- 函数顺序逻辑
- 数据段构造
- 漏洞利用
- 编译器特征
- 时间/地点/编码等

关键数据行为异常

基线

- 终端用户行为历史，如A部门用户每天平均访问220次关键数据
- 外发敏感数据行为历史，如用户、设备、时间、频率、和目的地等
- 内部业务系统和服务器敏感数据访问历史

异常侦测

- 超过正常访问敏感数据次数5倍以上
- 使用压缩软件RAR打包大量敏感数据
- 向USB设备中密集大量拷贝敏感数据
- 用户或设备频繁外发加密文件
- 从内部服务器下载大量表单等数据
- 大量访问恶意域名（DNS隐蔽信道点滴外传）

更多场景实例

- 恶意域名相关威胁
- Web drive-by
- 非授权特权账户访问
- 异常登录行为
- Tor /P2P
- 文件内网扩散

威胁情报推送

每天新增情报数量

- C&C域名：几万条
- 钓鱼网站：数十万个
- 木马家族变异样本：数万个
- 入侵手法TTP等

使用机器学习引擎

- 极低更新成本的DGA域名判定
- 域名与网页内容协同发现钓鱼
- 木马家族同源变种预测识别
- 异常行为分析

分布式前置机器学习引擎的实现

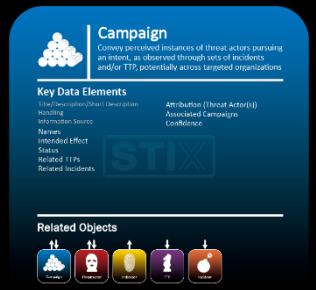
需求

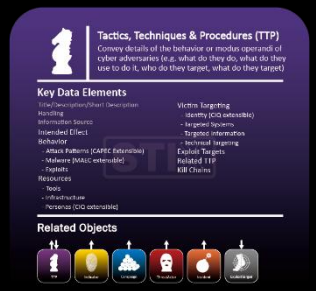
- 轻量化
- 场景针对性
- 高性能
- 产品级稳定
- 改进的响应速度

开源？太多工程难点

- 体积大
- 只有通用算法实现
- 实现性能难以接受
- 各种bug和功能变更
- 完全不可控

STIX速查卡 - 展台B6免费领取






ISC
2015
中国互联网安全大会
360互联网安全中心
cirrus
gate
思睿嘉得