大数据中的安全解决之道

——大数据助力网络安全

李 鑫

Lixin011@huawei.com

目录

- Big Data, Big Security Problems
- 方法与实践
- 观点分享

Big Data

互联网用户：27亿

移动宽带用户：20亿

--国际电信联盟

每天新增恶意软件：20,000

--卡巴斯基

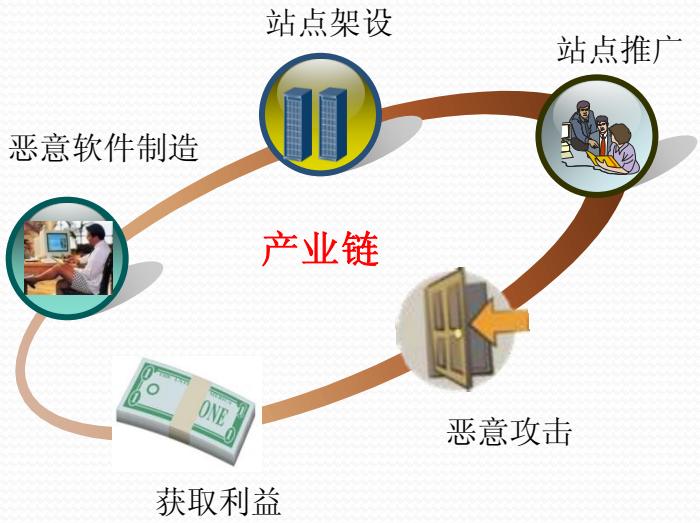
网络犯罪受害者人数：5.56亿

网络犯罪造成的净损失：$1,100亿

--赛门铁克

Big Security Problems

- 恶意数据混杂在正常数据中
- 恶意软件制造专业化
- 很多逃避检测的方法
- 恶意软件生存期短



目录

- Big Data, Big Security Problems
- 方法与实践
- 观点分享

安全检测与大数据融合

收集

- ✓可执行文件
- ✓图片
- ✓压缩包
- ✓页面
- ✓流量
- ✓……

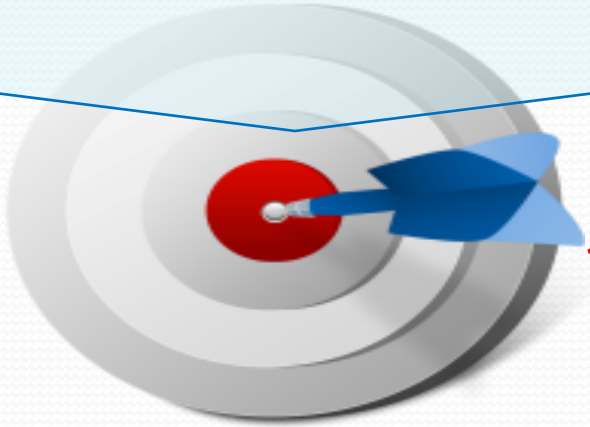
提炼

- ➢分类
- ➢关联分析
- ➢数据挖掘

检测

- ➢安全分析
- ➢数据挖掘

长尾数据



发现威胁

数据提炼—千万级数据的多维度分类

内容

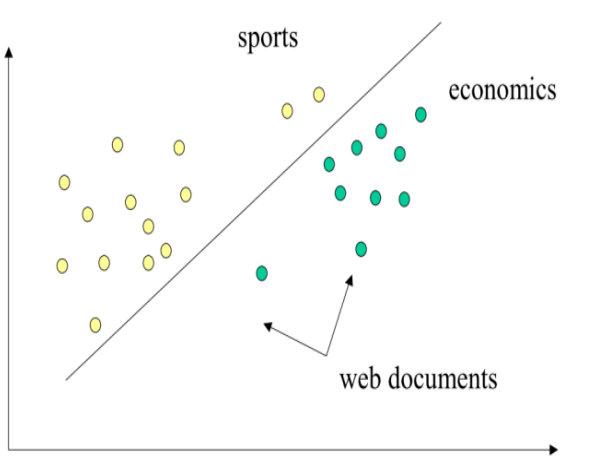
- ◆黄、赌、毒
- ◆金融理财
- ◆竞技体育

功能

- ◆bbs/论坛等交互性站点
- ◆门户、搜索、Mail
- ◆代理服务器

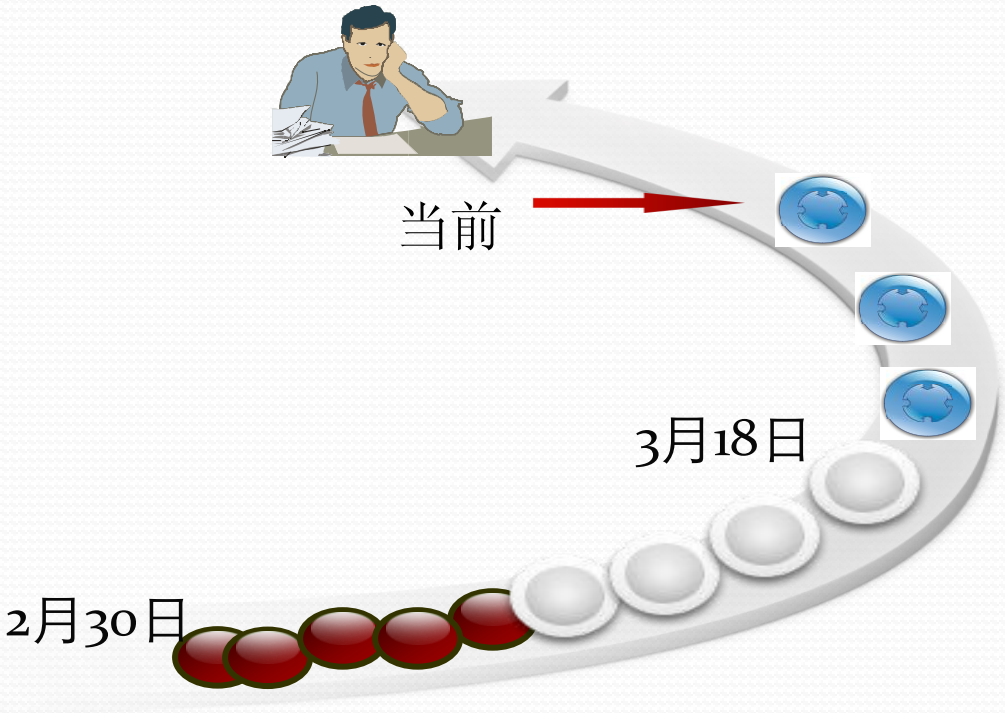
平台

- ◆操作系统
- ◆平台软件



数据提炼—亿级URL/IP数据的生命周期评估

- 数据检测历史结果
- 数据的当前检测结果

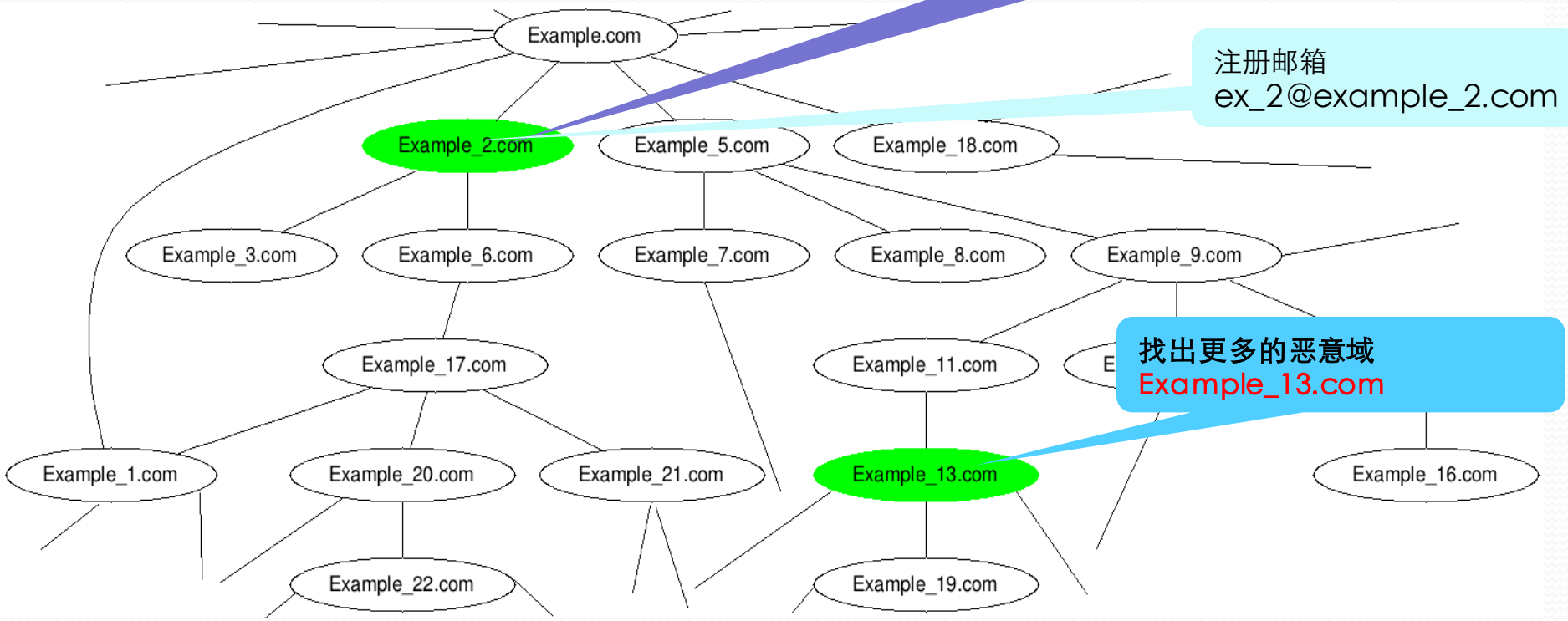


关联分析—数化的

已知的恶意域
example_2.com

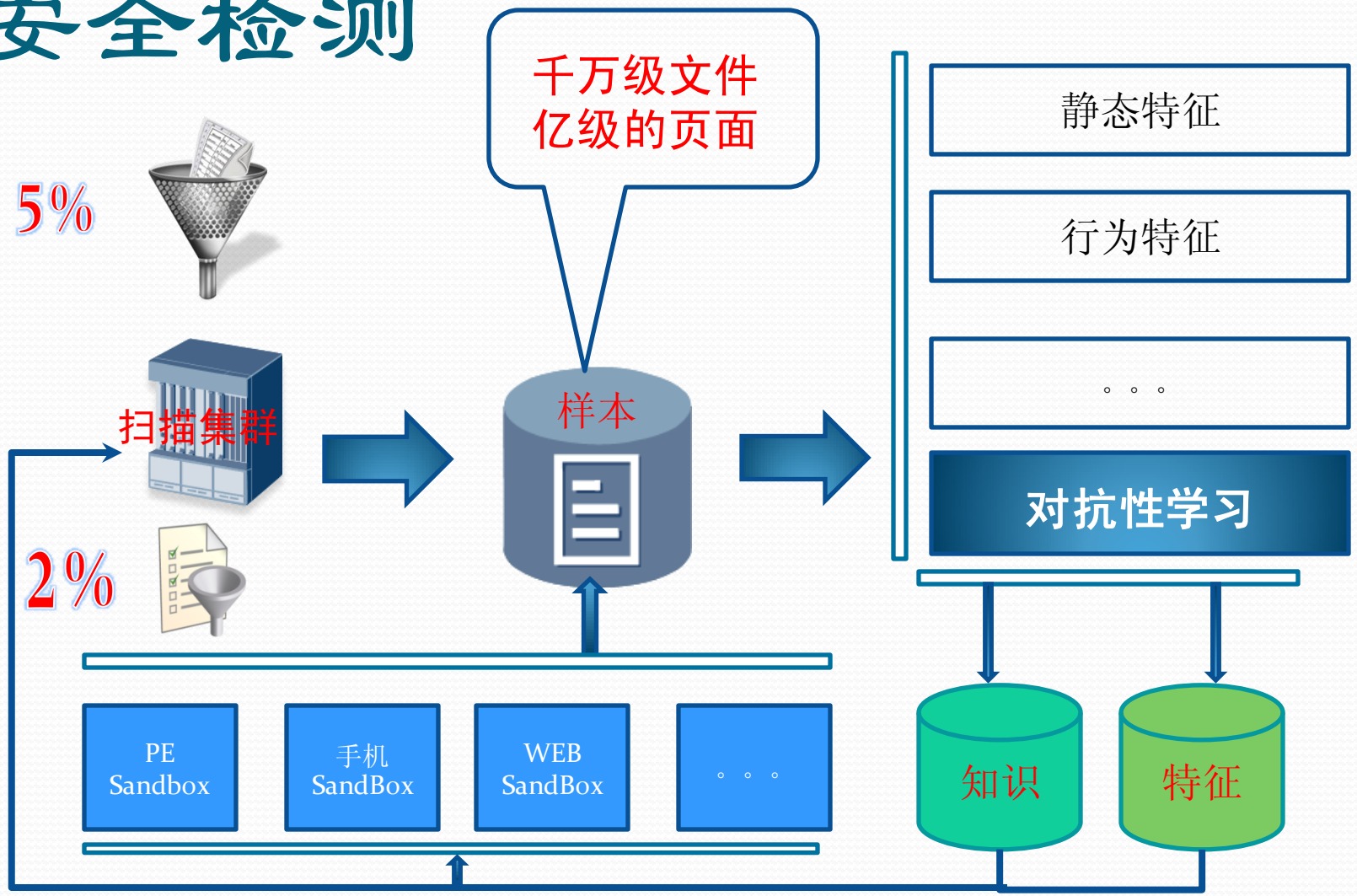
注册邮箱
ex_2@example_2.com

找出更多的恶意域
Example_13.com

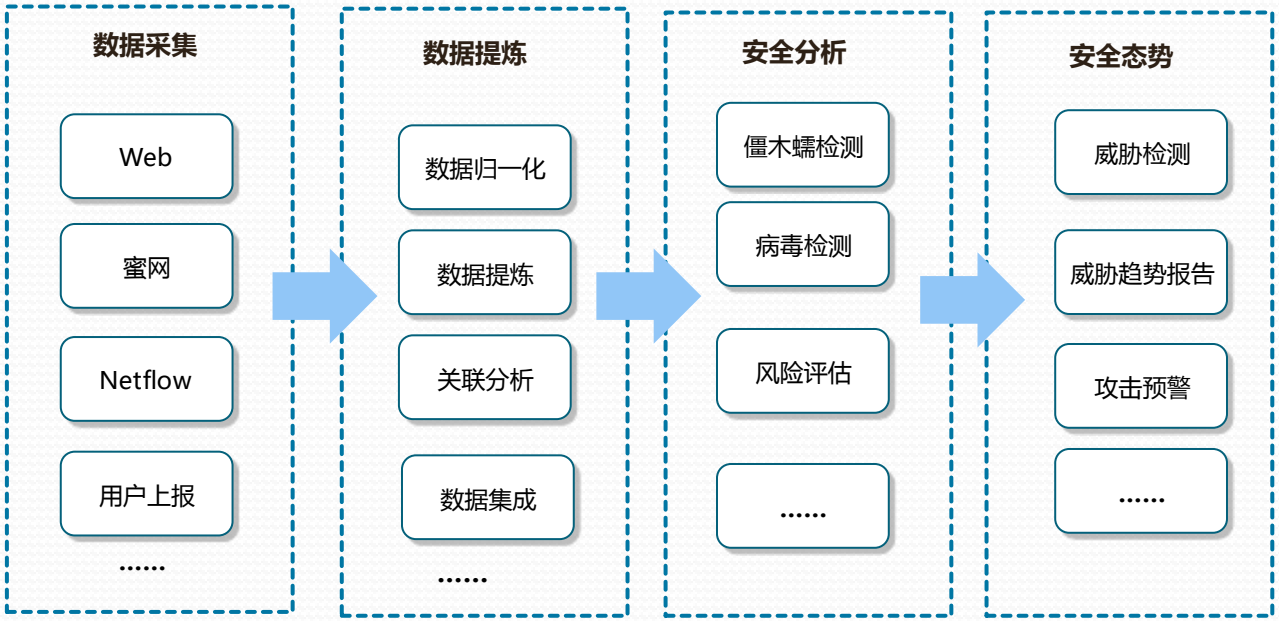


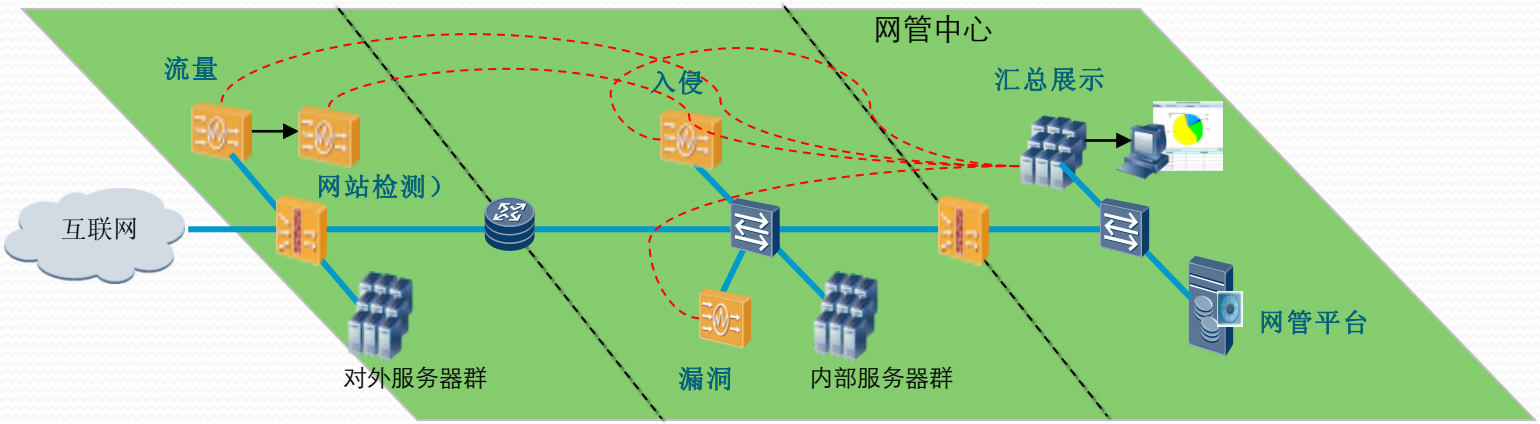
邮箱关联 + 域名关联 + IP关联 + … = 结果

安全检测


千万级文件
亿级的页面
5%
扫描集群
2%
样本
静态特征
行为特征
。。。
对抗性学习
PE
Sandbox
手机
SandBox
WEB
SandBox
。。。
知识
特征

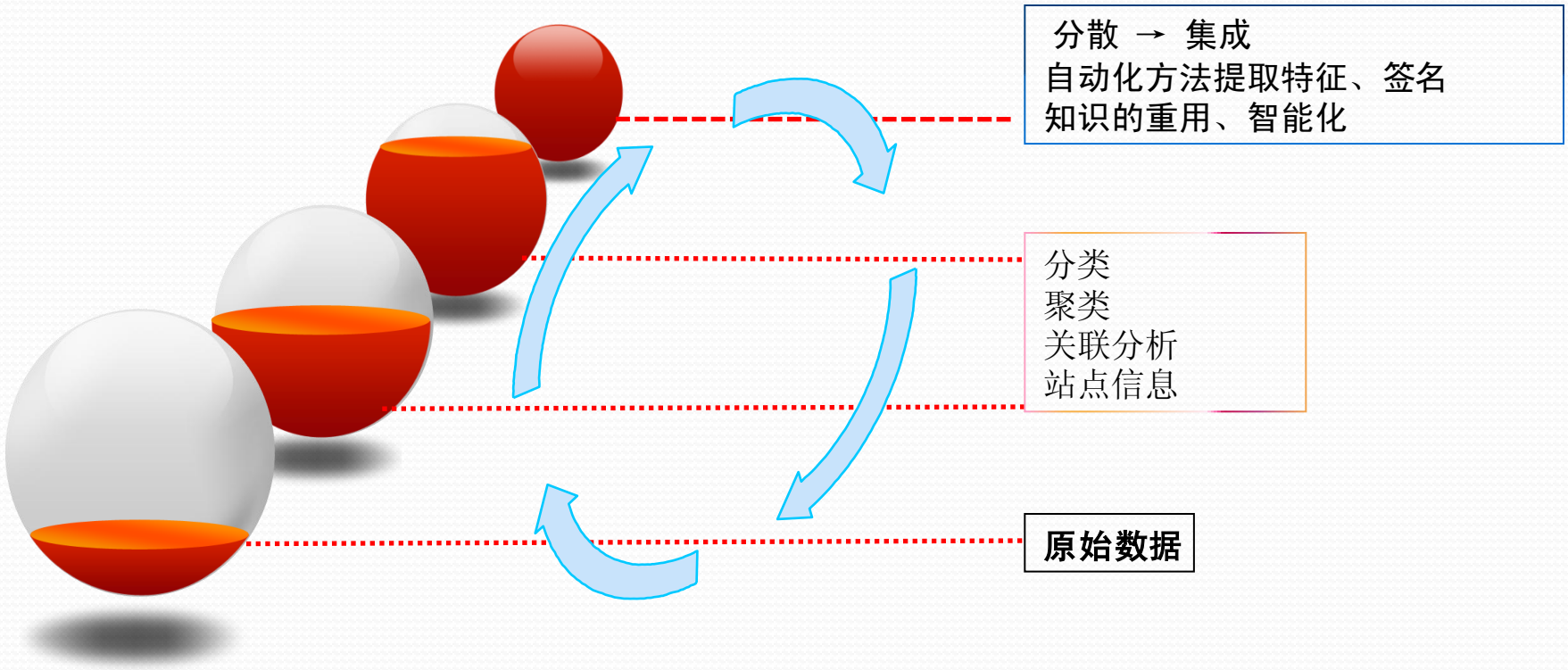
安全智能中心—每天千万级数据的分析





安全分析模式变化

- 海量的待分析数据与相对有限的分析能力的矛盾



目录

- Big Data, Big Security Problems
- 方法与实践
- **观点分享**

- 大数据，更多的安全问题；更丰富的安全视角
- 单个样本分析 -> 多个样本关联 -> 海量样本挖掘
- 结合数据上下文的安全分析
- 基于对抗性的机器学习、关联分析

谢 谢