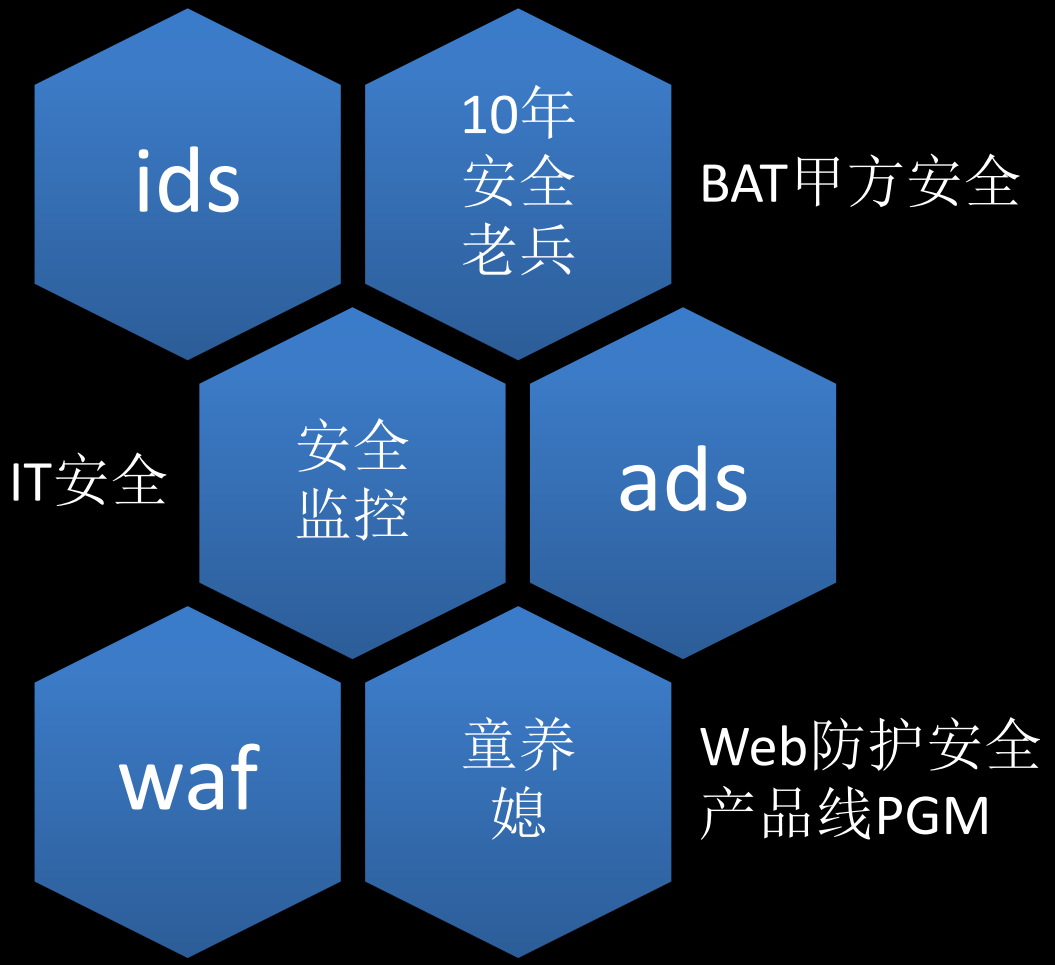
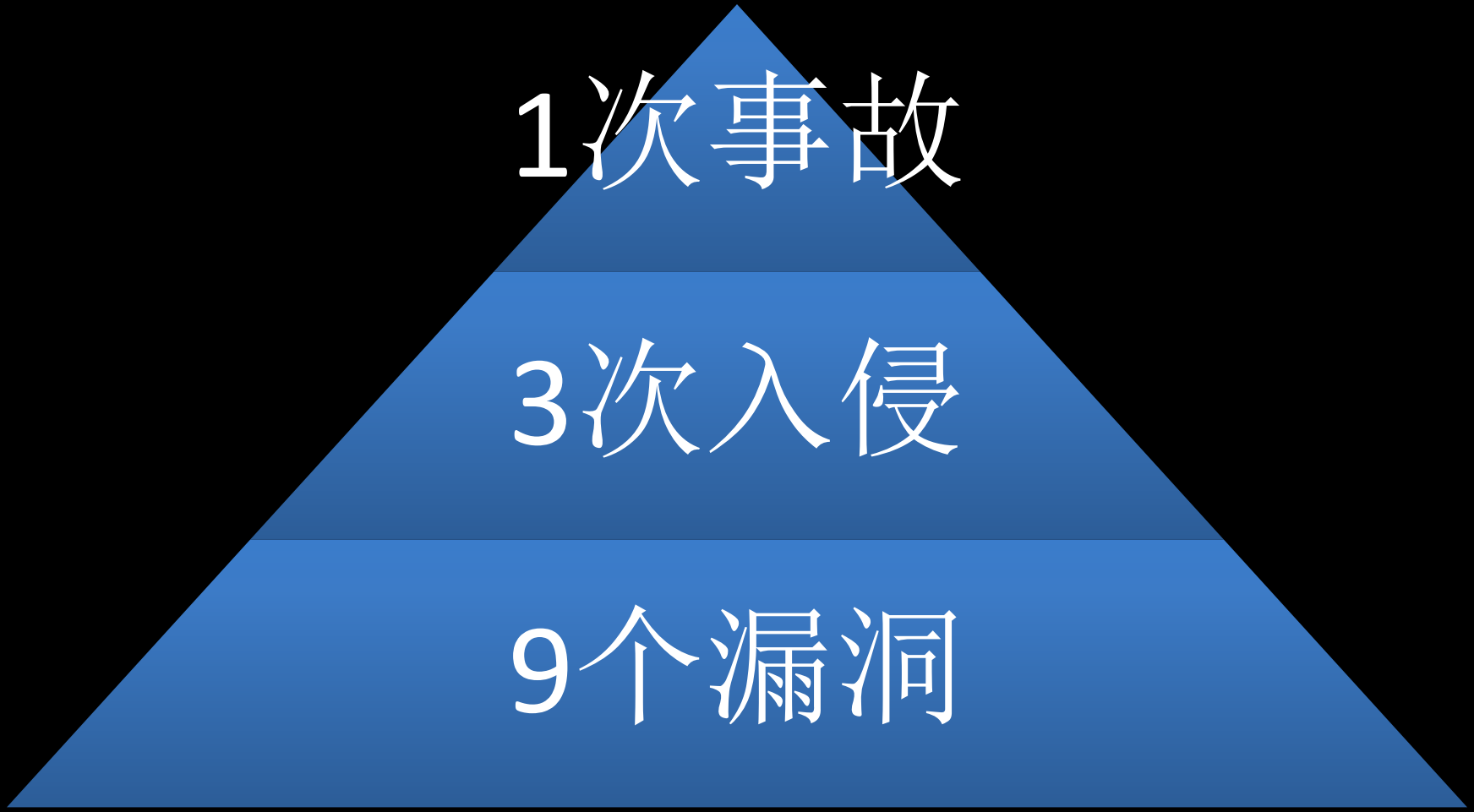
企业安全监控经验教训分享

百度安全 刘焱

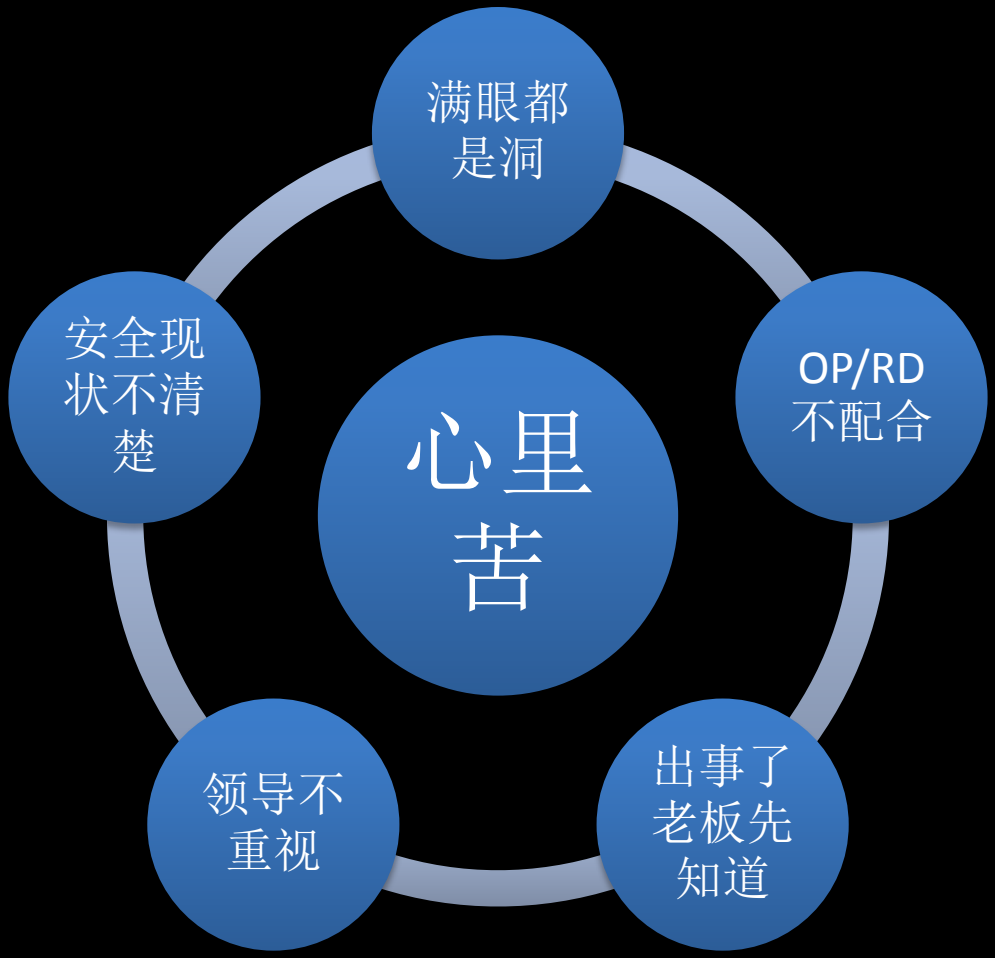
自我介绍



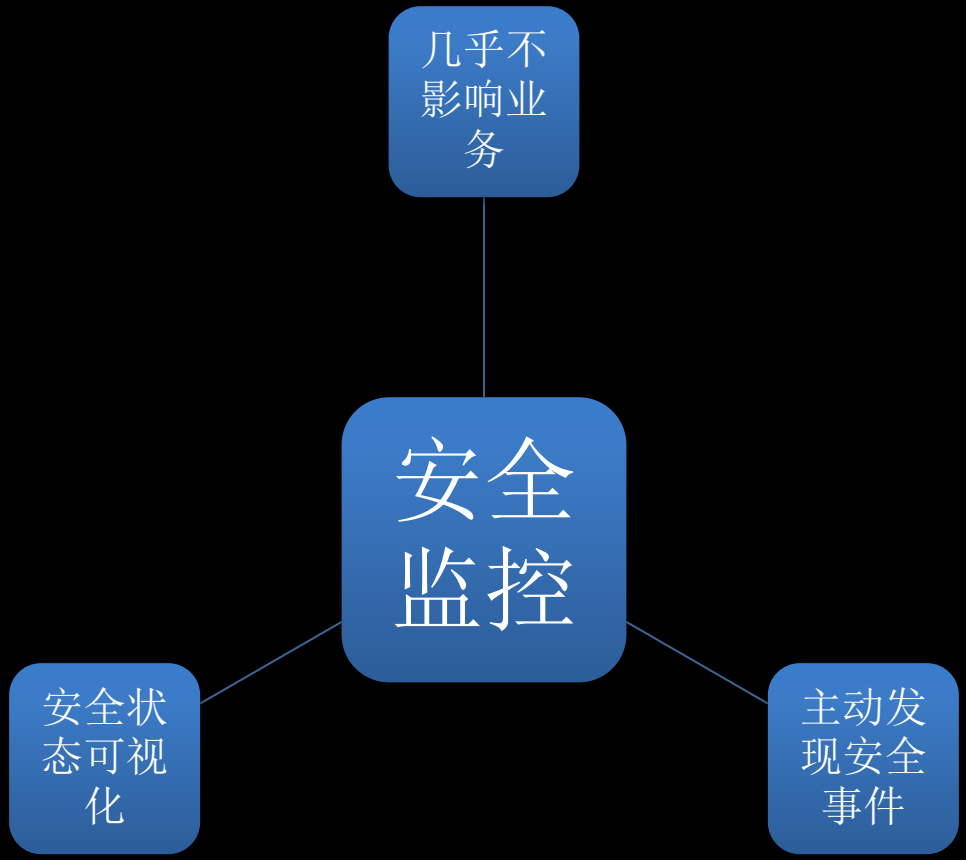
出来混 总是要还的



甲方安全不好搞



安全建设 监控先行



百度安全监控发展历程

基于OSSIM，linux服务器、邮件、IPS、准入、VPN、全流量的统一分析处理

基于hadoop的数据库、web服务器日志分析

基于自研安全客户端的webshell检测

基于storm/spark、ELK体系的实时报警处理

机器学习、威胁情报初步应用

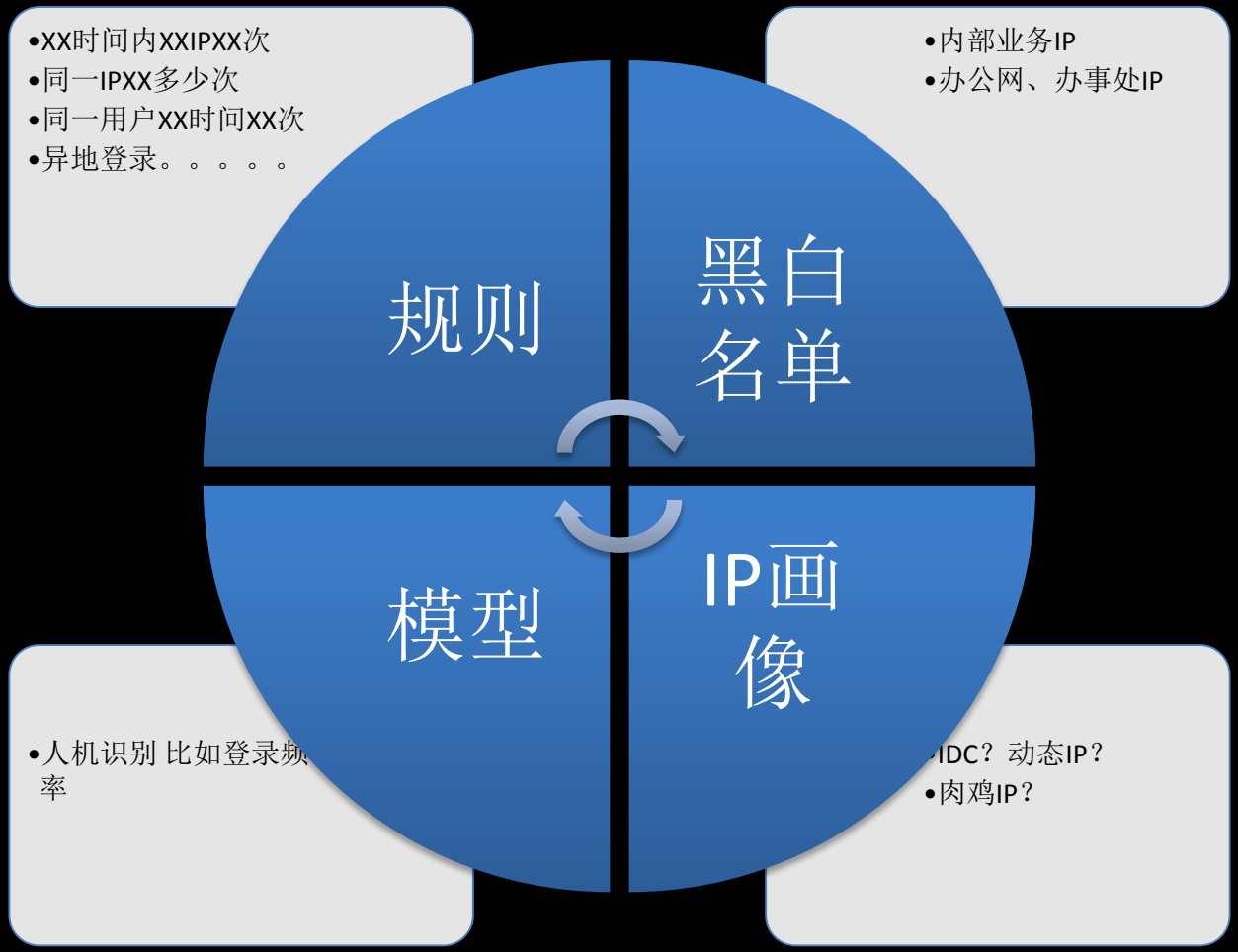
邮箱撞库、拖库、后门

邮箱撞库	• 财报、预算、人事薪酬等核心运营数据 • 需求、设计、产品规划等核心研发信息 • 重要系统的密码
拖库	• 用户密码、个人信息 • 交易、支付、订单信息 • 真实DAU
后门	• 主页篡改 • 部署恶意脚本搜集用户信息 • 源码泄露

在合适的地方做合适的事

	第一优先级	第二优先级
邮箱撞库	邮箱日志分析	
拖库	数据库查询日志分析	基于web服务器流量分析 基于数据库服务器文件、流量监控
后门	Web服务器文件分析 Linux服务器bash反弹shell监控	基于web服务器流量分析

邮箱撞库监控-经验


•XX时间内XXIPXX次
•同一IPXX多少次
•同一用户XX时间XX次
•异地登录。。。。。
规则
黑白名单
•内部业务IP
•办公网、办事处IP
模型
IP画像
•人机识别 比如登录频率
IDC？动态IP？
•肉鸡IP？

邮箱撞库监控-坑点

- 微软邮件服务器角色众多，令人费解、edge、cas、nlb啥的，现在还是一头雾水
- 默认配置下部分服务器上看不到客户端IP，it折腾半天搞好了，也不说咋回事
- 手机把异地登录搞得很复杂，见过天津手机到北京，出口IP还是天津的。。。

拖库检测-经验

数据库日志

- 与DBA合作，分析处理了百度DBA托管的全部数据库日志，覆盖凤巢、大搜、糯米，日均处理4T，日志从生成到处理秒级延时，spark+storm
- 规则+SQL沙盒
- 乌云&BSRC SQL注入主动发现80%以上

全流量

- 覆盖百度部分重点机房
- 规则+SQL沙盒
- 乌云&BSRC SQL注入主动发现80%以上
- 重点解决测试开发区域以及小产品线未进入DBA托管的

拖库检测-坑点

- 数据库日志搜集之前是在客户端，性能消耗较大，还好DBA支持。。。
- 数据脱敏，大家都放心点。。。
- 越是开发测试区域越容易出事
- 神一般的RD自运维，各种奇葩备份脚本
- 见过8M大小的SQL语句，不像人写出来的，但是确实是人写的。。。

后门检测-经验

Webdir1.0

- 覆盖百度全部对外发布的web服务器
- 基于notify机制秒级检测

Webdir2.0

- 覆盖百度全部web服务器，百万级部署，支持docker等环境
- 规则+沙盒
- 获得百度创新奖最高级别
- 乌云&BSRC后门主动发现率80%

Webdir3.0

- 流量与webdir联动实现了精准报警，邮件、短信、电话、、、
- 机器学习初步尝试，确实可以发现很诡异的冬冬

后门检测-坑点

- 神一般的rd，写的各类脚本堪比大马
- 线上环境大量执行代码include外部php代码
- PHP语法过于灵活，规则实在搞不定才想起搞沙盒
- 各类管理平台、管理后台容易造成机器学习算法的误报

Q&A