

AI在数据安全中的实践





彭思翔 博士

腾讯专家研究员

目录

- 01 新时代的数据安全挑战
- 02 操作审计：基于AI的数据库审计
- 03 隐私保护：数据分析/共享中的隐私保护
- 04 数据防泄密：异常用户行为分析



新时代的数据安全挑战



隐私保护法规趋严



中国:等级保护 2.0时代



欧盟: GDPR 2018



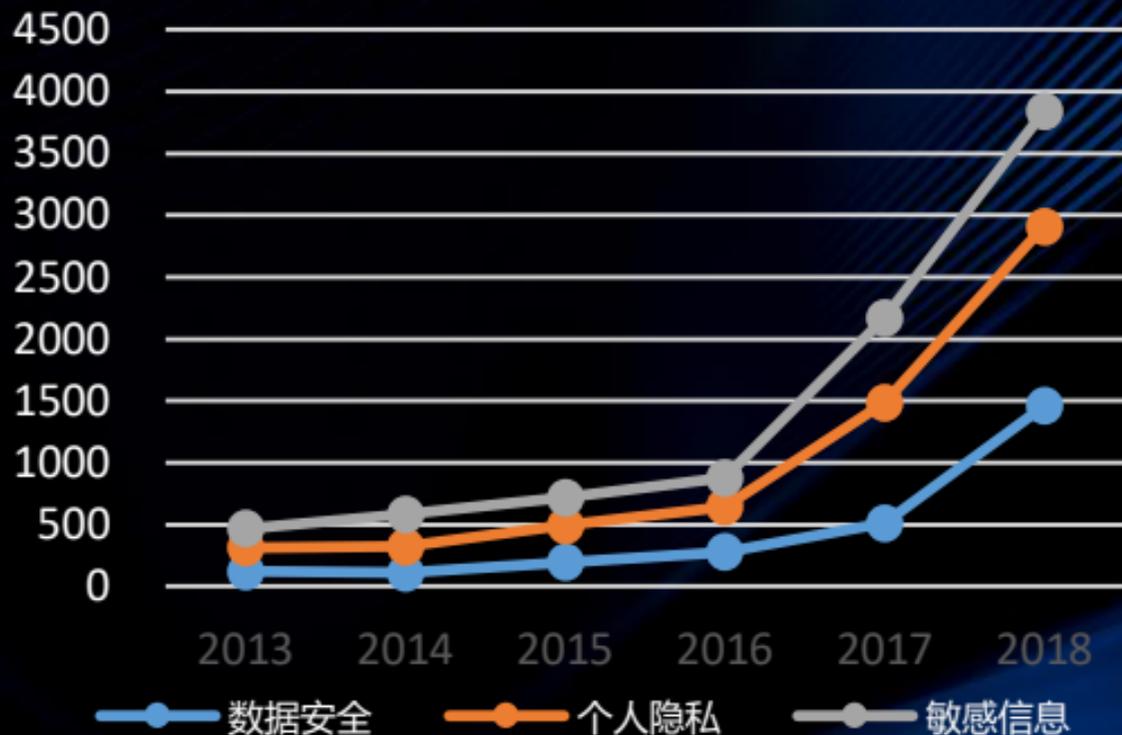
美国: TrustArc 20+

个人隐私保护意识觉醒

隐私保护受关注度

年增长 **80%**

互联网上相关主题数量 (单位: 万)



数据安全态势严重

- ◆ 内部威胁通常很难发现，一般都在2个月以上
- ◆ 很难将有害行为与正常工作区分开来



◆ Crowd Research Partners 发布的《2017年企业内部威胁报告》指出，一个组织每年遭遇内部威胁的平均成本超过800万美元。

- ◆ 金雅拓《全球公共数据泄露水平指数》2017年统计，全球公开的数据泄露事件导致26亿数据泄露
- ◆ 恶意内部事件泄密量同比增长4,114%

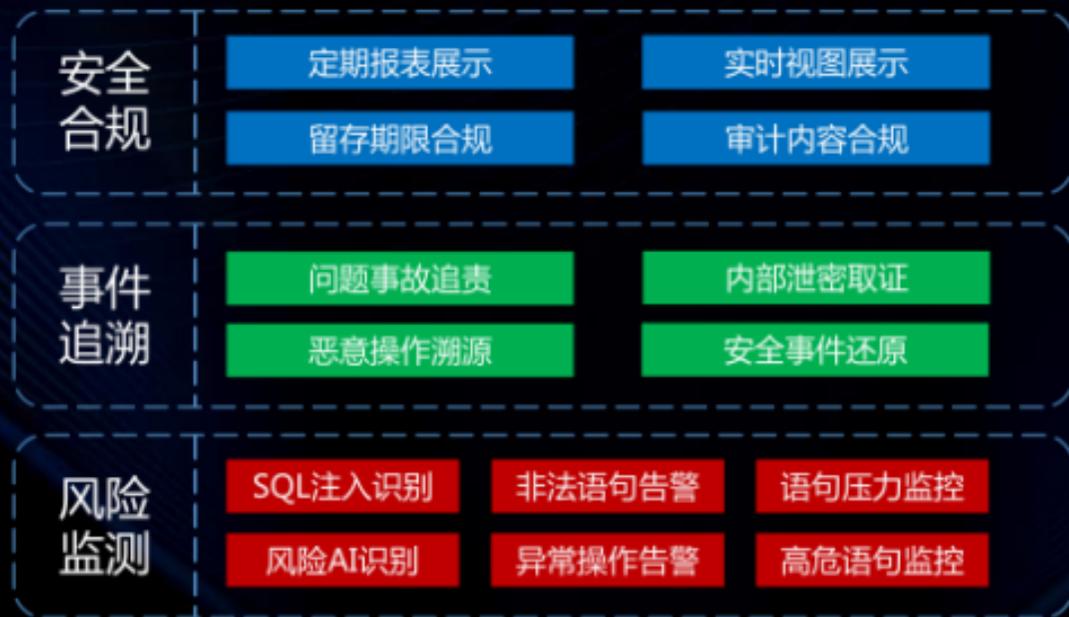
腾讯安全数盾：全景图





操作审计：基于AI的数据库审计

操作审计：合规的基石



等保合规



事故定责



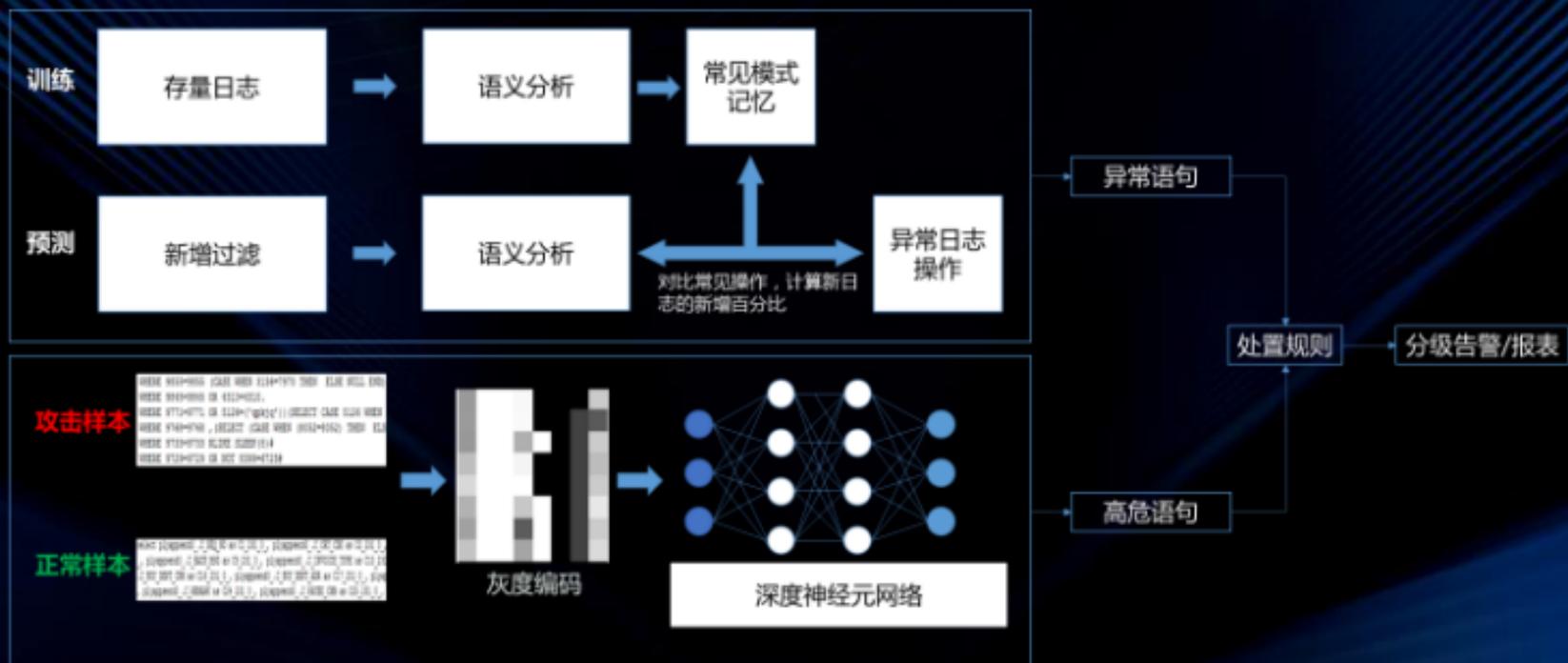
防统方



泄密取证

操作审计：AI赋能，精准发现

独创AI双引擎综合判断，自动适配用户操作特征，覆盖率:99%,误报率百万分之一

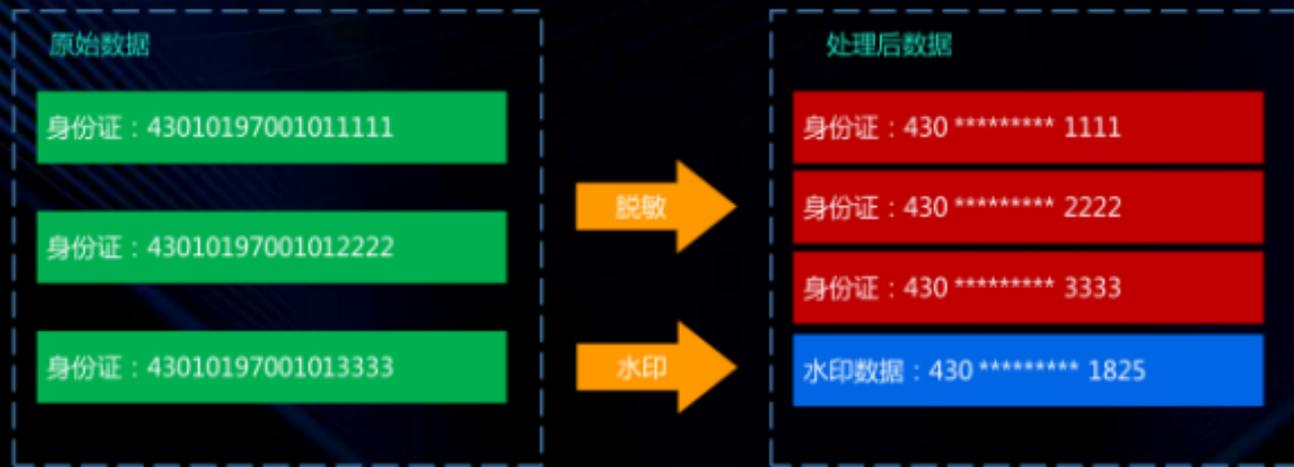




隐私保护：数据分析/共享中的敏感
信息特殊处理

隐私保护：数据脱敏、水印

对敏感数据进行脱敏和水印处理，同时保持数据统计学价值
满足数据分析/共享环境中的隐私保护需求



系统测试



数据分析



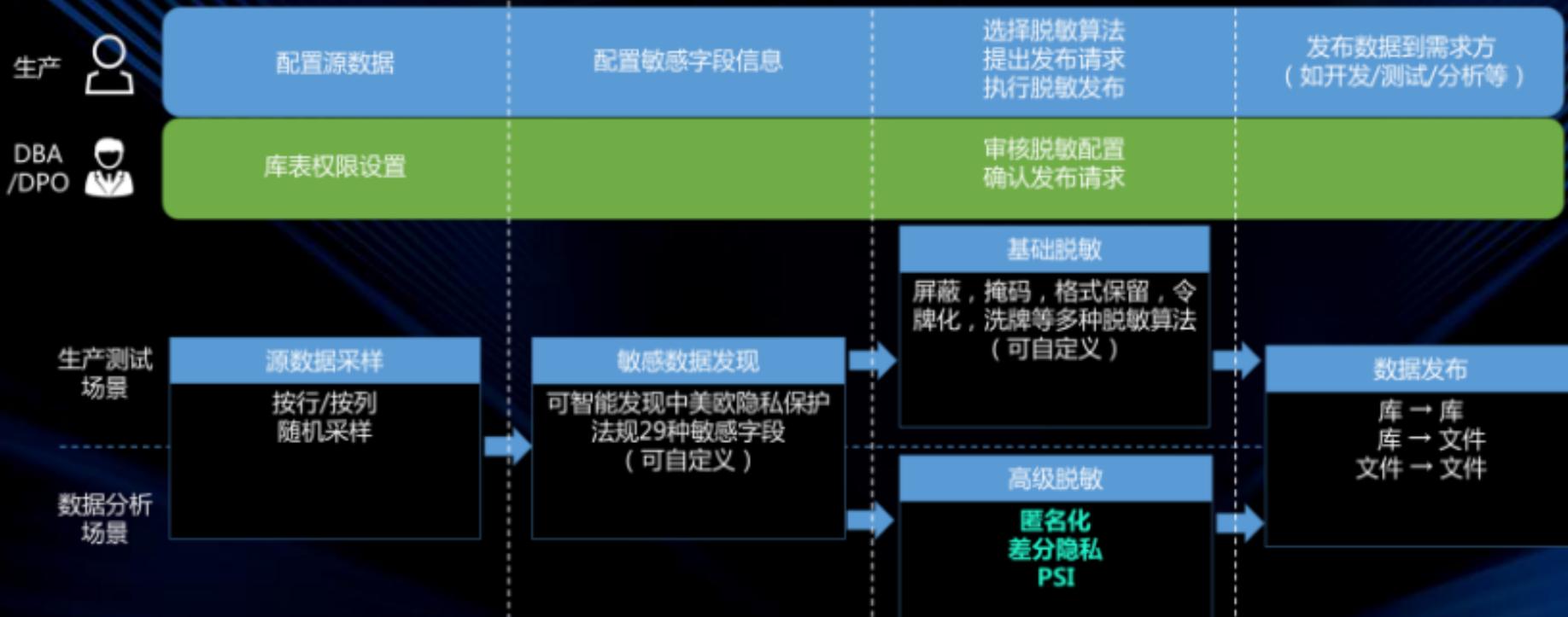
应用开发



业务培训

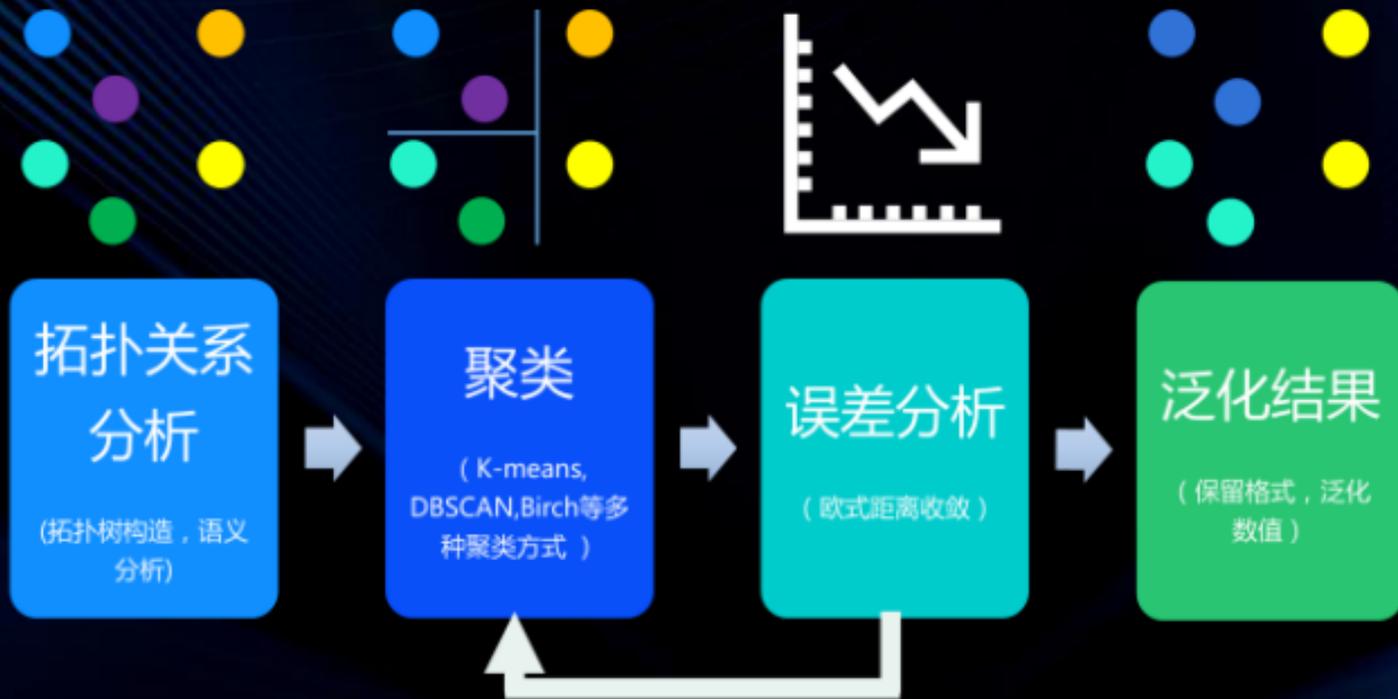
隐私保护：智能高效

一键智能脱敏，满足生产数据用于测试、开发、培训和大数据分析场景中的数据脱敏需求



隐私保护：平衡隐私保护与数据挖掘价值

匿名化：通过聚合与泛化使得脱敏后数据无法被唯一对应，同时保证统计分析可用性
适合离线数据批量脱敏，支持多种数据类型，各种业务通用



隐私保护：平衡隐私保护与数据挖掘价值

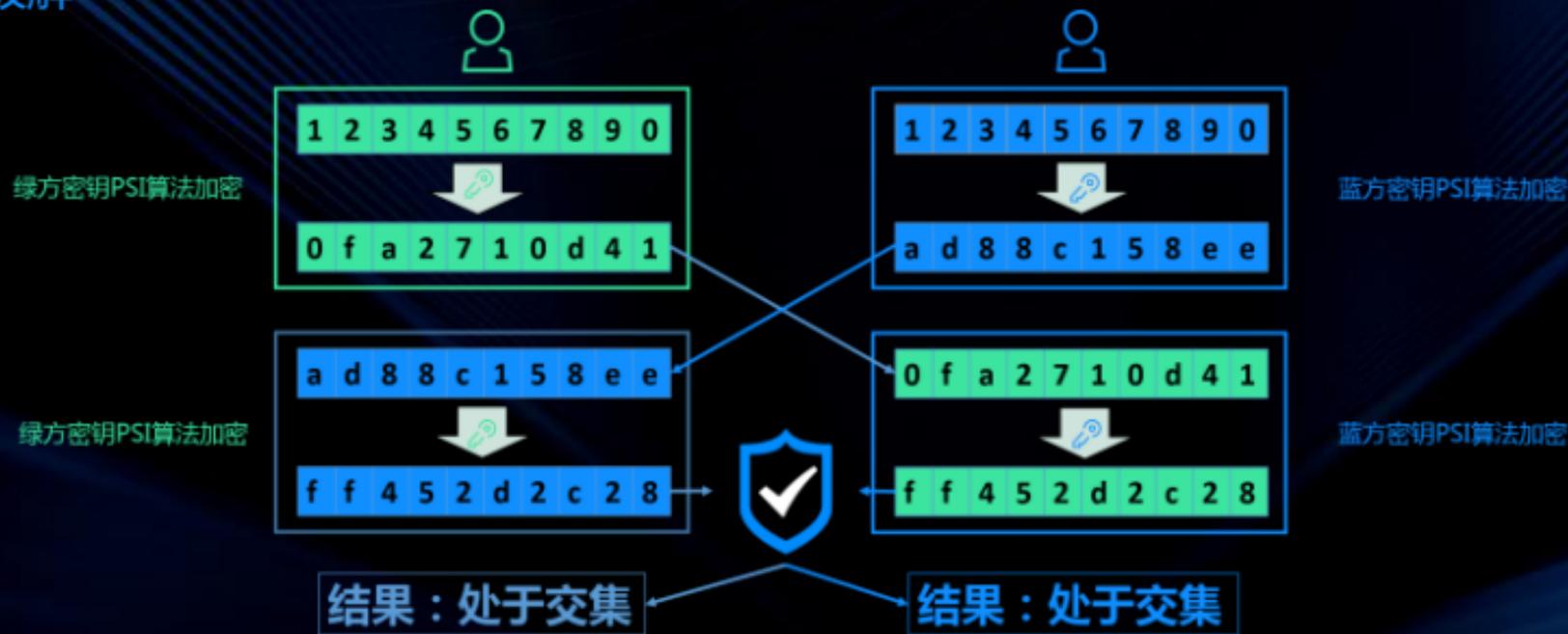
差分隐私：通过语义分析，对结果加入噪声，平衡隐私安全与数据可用性
适合实时数据查询结果脱敏，具备高保护等级，低数据分析误差



隐私保护：平衡隐私保护与数据挖掘价值

安全多方计算框架：针对多方互不信任但需要共享数据的场景，解决安全分析问题

数隐采用Private Set Intersection(PSI) 共享安全算法，各方地位平等，线性计算复杂度，无法被碰撞破解





数据防泄密：异常用户行为分析



数据防泄漏：攻击手段多种多样

基于规则的防护体系已经不能解决内部泄漏威胁

管理手段

账户权限管控

源IP段限制

命令限制防拖库

文件外发审计

全量日志审计

改密策略

应对策略

恶意提权操作

内部盗号+跳板攻击

定时脚本每天少量下载，积累核心数据

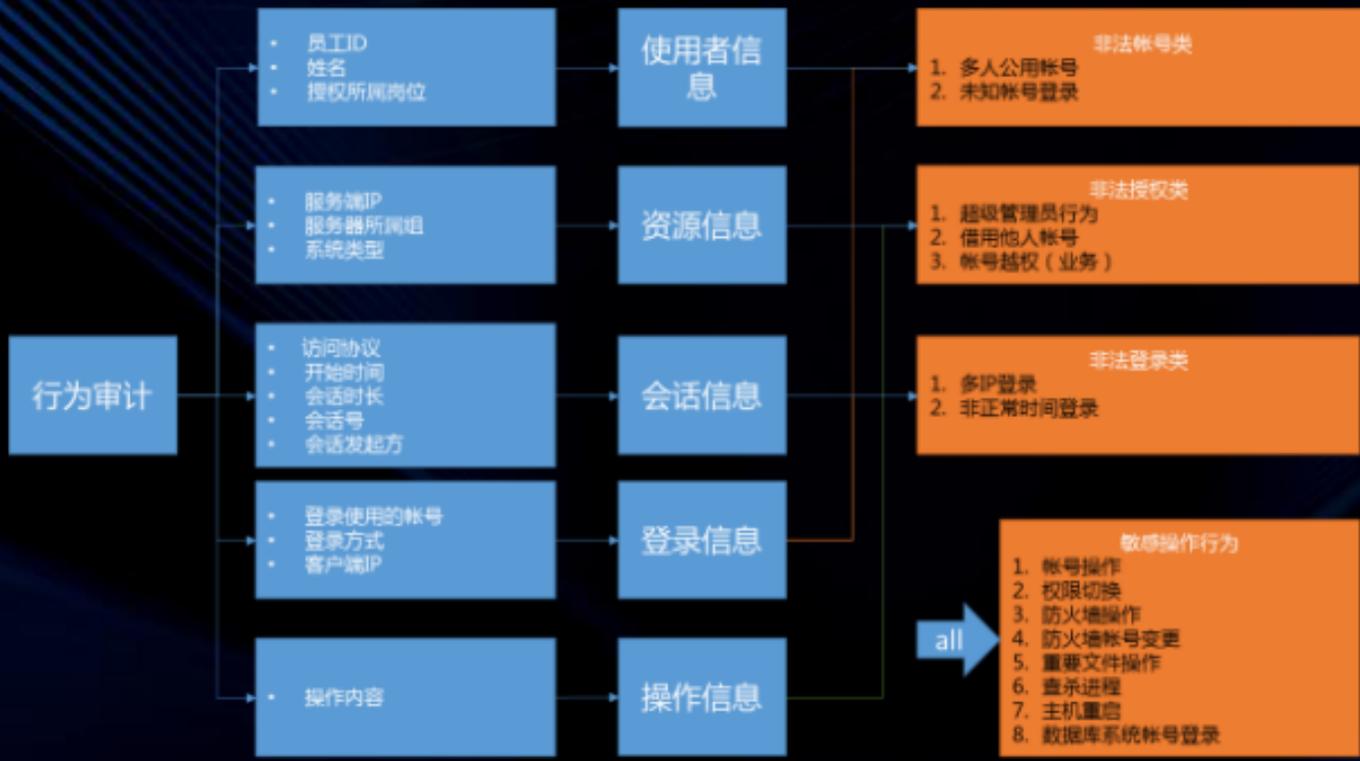
隐藏脚本控制服务器自行外发数据

数据泄密不直接影响业务连续性，泄密后长时间无人追溯

低频暴力破解

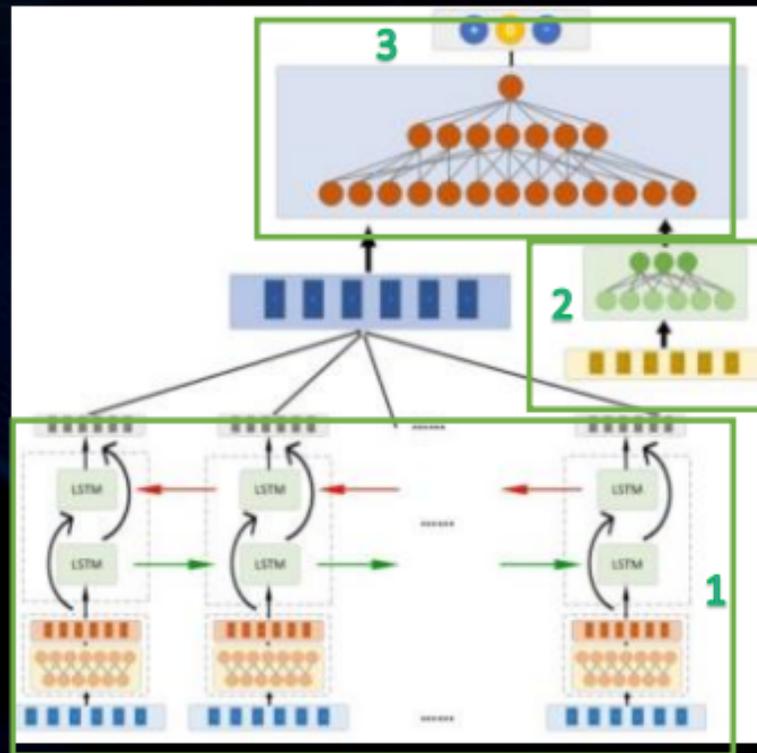
数据防泄漏：用户异常行为分析

使用数据安全网关收口对敏感数据资产的操作，并进行异常行为分析



数据防泄漏：用户异常行为分析

使用LSTM+CNN混合神经网络建立用户行为基线为并发现异常



当期用户敏感操作行为预测

用户历史行为画像：
敏感操作行为统计

用户当期操作行为时序：
ID,登录信息, 操作信息





THANKS

— TENCENT SECURITY CONFERENCE 2019 —