携程业务安全前生今世

携程技术保障中心高级安全工程师 闵杰

2016携程信息安全沙龙

个人介绍

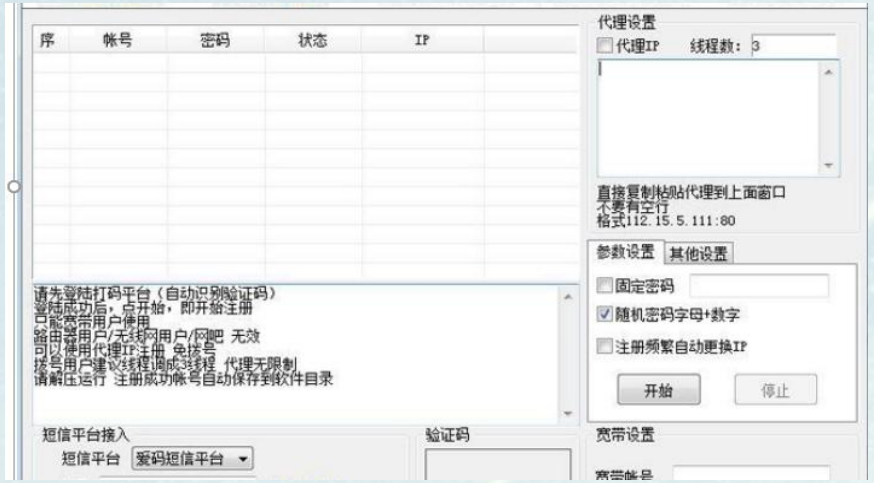
账户安全工程师

信息安全工程师

目录 Contents

业务安全问题

2016携程信息安全沙龙



垃圾注册：

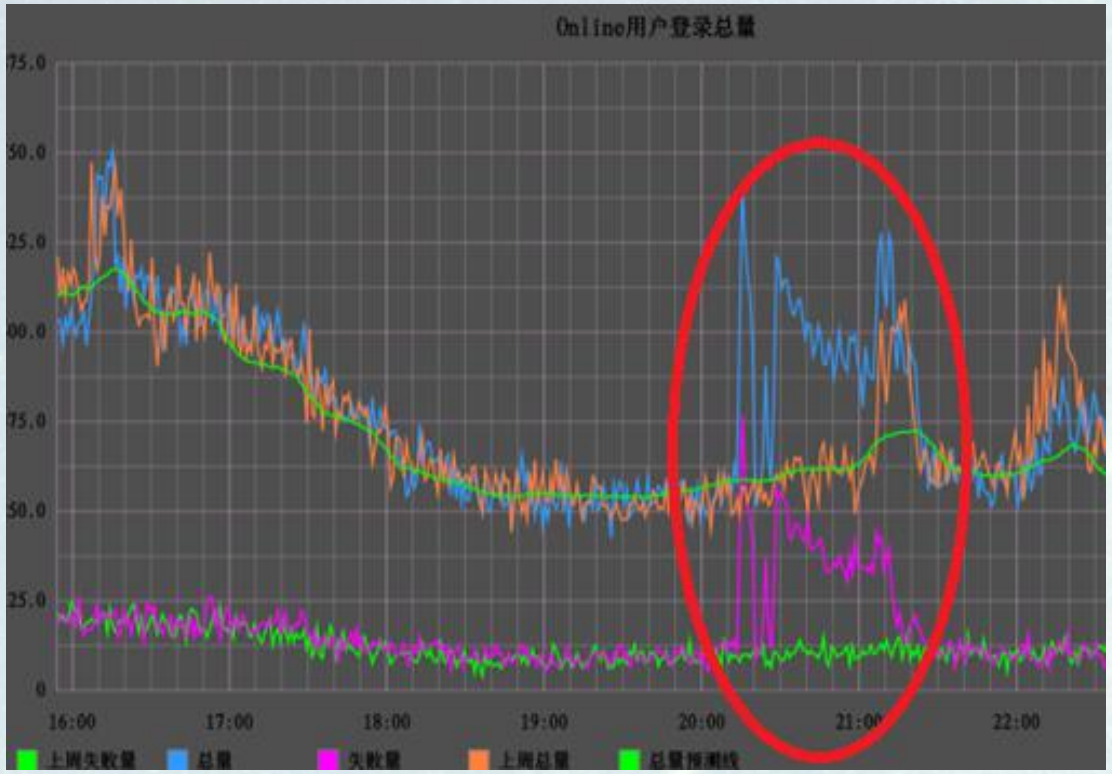
业务风险的源头之一

难点：

注册IP和设备指纹分散

识别验证码

注册渠道多样化



扫号：

威胁账号安全的关键点

资金盗用

信息泄漏

恶意欺诈

难点：

IP使用量巨大，可以做到1号1IP

使用外部社工库，密码正确率高

可以根据安全措施及时更换策略

设备指纹基本伪造，无明显特征

薅羊毛：

影响活动实际收益和到达率

侵占有限的活动资源

难点：

牟利方式多样化，各种形式组合

模拟真人或直接真人操作

黑色产业链发达，集团化模式

原 【机器人系列】爬取携程产品图片式价格

标签： 爬虫 价格破解 数字识别

2015-03-26 18:05 1215人阅读 评论(5) 收藏 举报

分类：

JAVA Web编程（9）

携程旅行网是国内最大的在线旅游提供商，其价格为了防止爬虫，是用了图片形式，从而防爬。 据我所“爬”，美团最近也开始使用图片形式的价格。但是这种图片说白了其实是自欺欺人，防君子不防小人（应该是防菜鸟不防高手才对）。今天，咱们就来看看，如何破解携程的图片式价格。

先上一张图，看看这个价格是怎么来的。

可以看到，这个数字5，是由p_h57_7这个CSS样式定义的。而这个样式里定义了一个背景图片，注意这个地方后面跟了一个数字！ 也就是 -1346。 看看这个图片是啥样的~

爬虫：

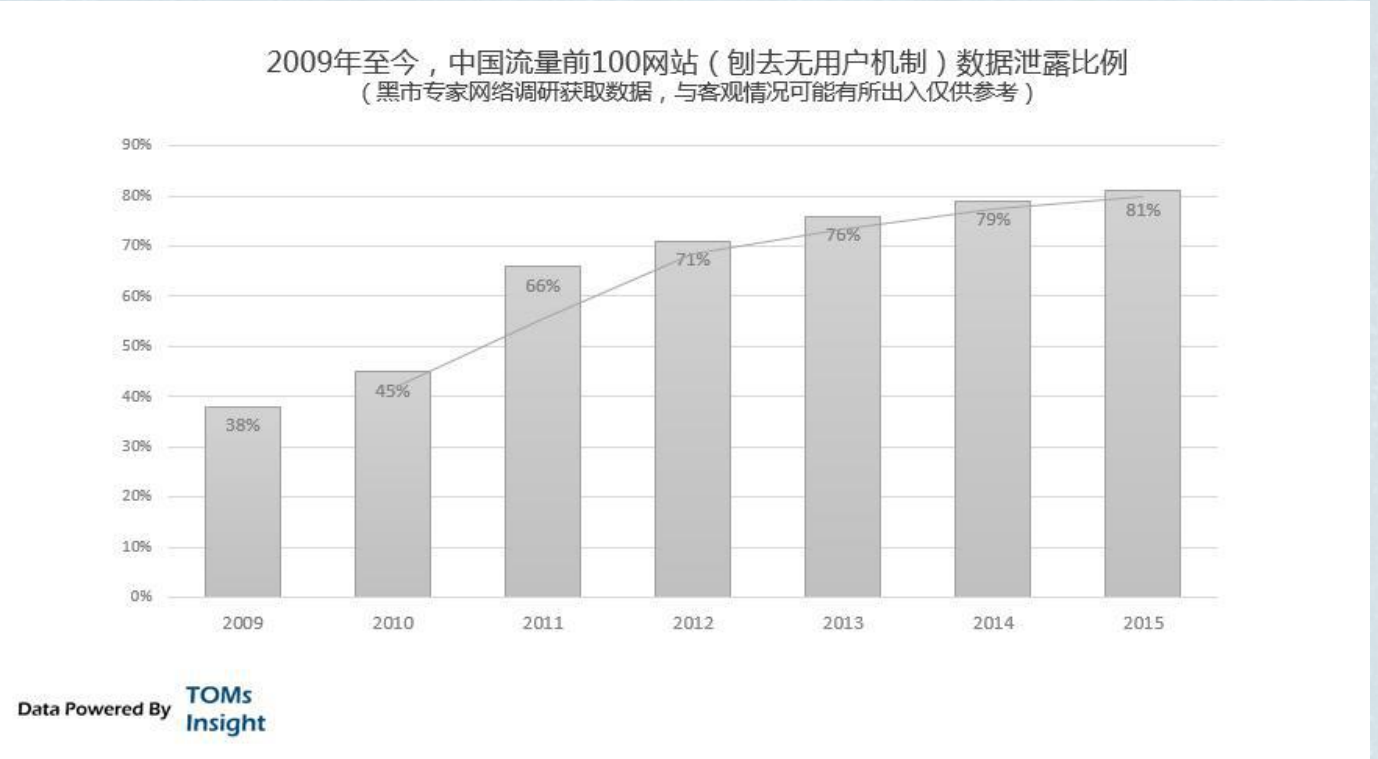
企业的价格策略被掌握

扰乱PV/UV,无法做出正确营销判断

难点：

频次低，特征不明显

不会对业务方造成明显感知



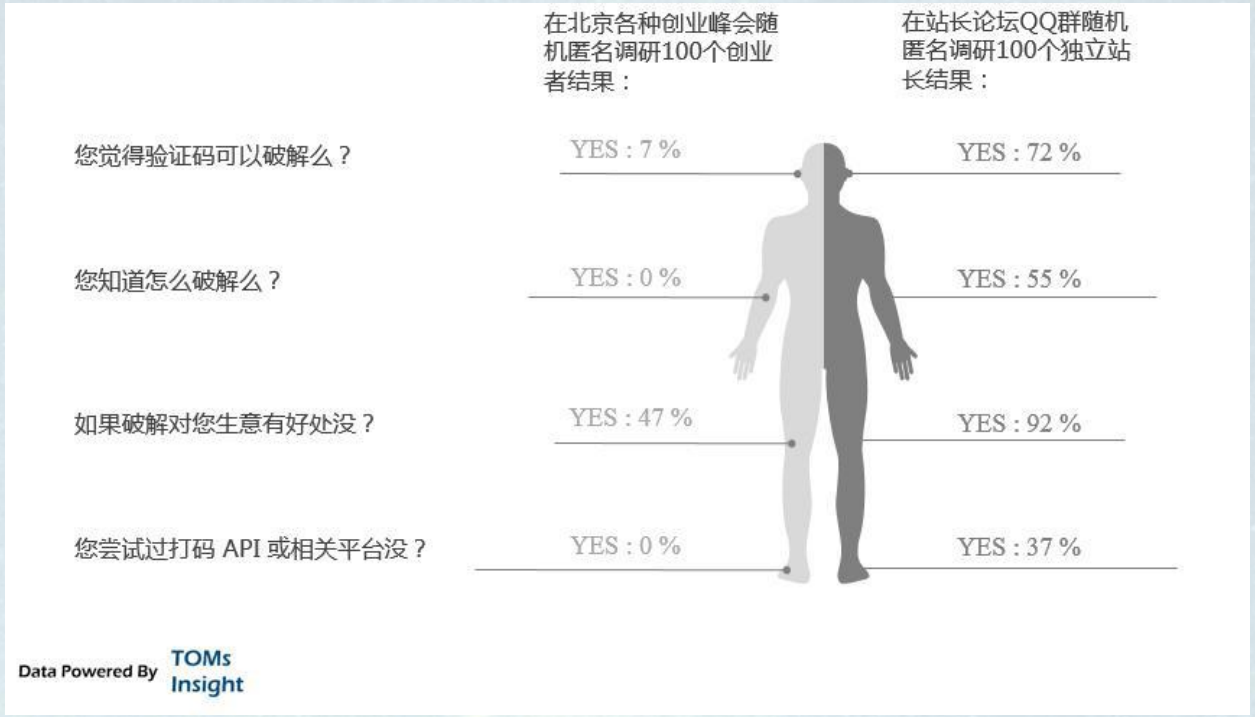
第三方调查结果：

实际上大部分的脱库事件和暴露在公众面前的数据泄漏，对真正的地下产业链并没有多少价值，原因是接近80%的数据都已经被掌握了。关键词：十墓九空

第三方调查结果：

打码平台工作室日益增多，现在很多的批量注册和撞库工具都内置集成了打码平台模块，形成了一个新兴的产业链。

目前国内相关的从业人员已经达到了百万级别。



第三方调查结果：

验证码破解方法以及收益已经被超过半数以上的草根互联网人所知晓，并且有超过1/3的草根站长尝试过打码平台。这股力量会成为以后甲方业务安全的重要阻力。

后援保障现状

指纹信息覆盖低

数据纬度少

沟通成本大

账户安全设计陈旧

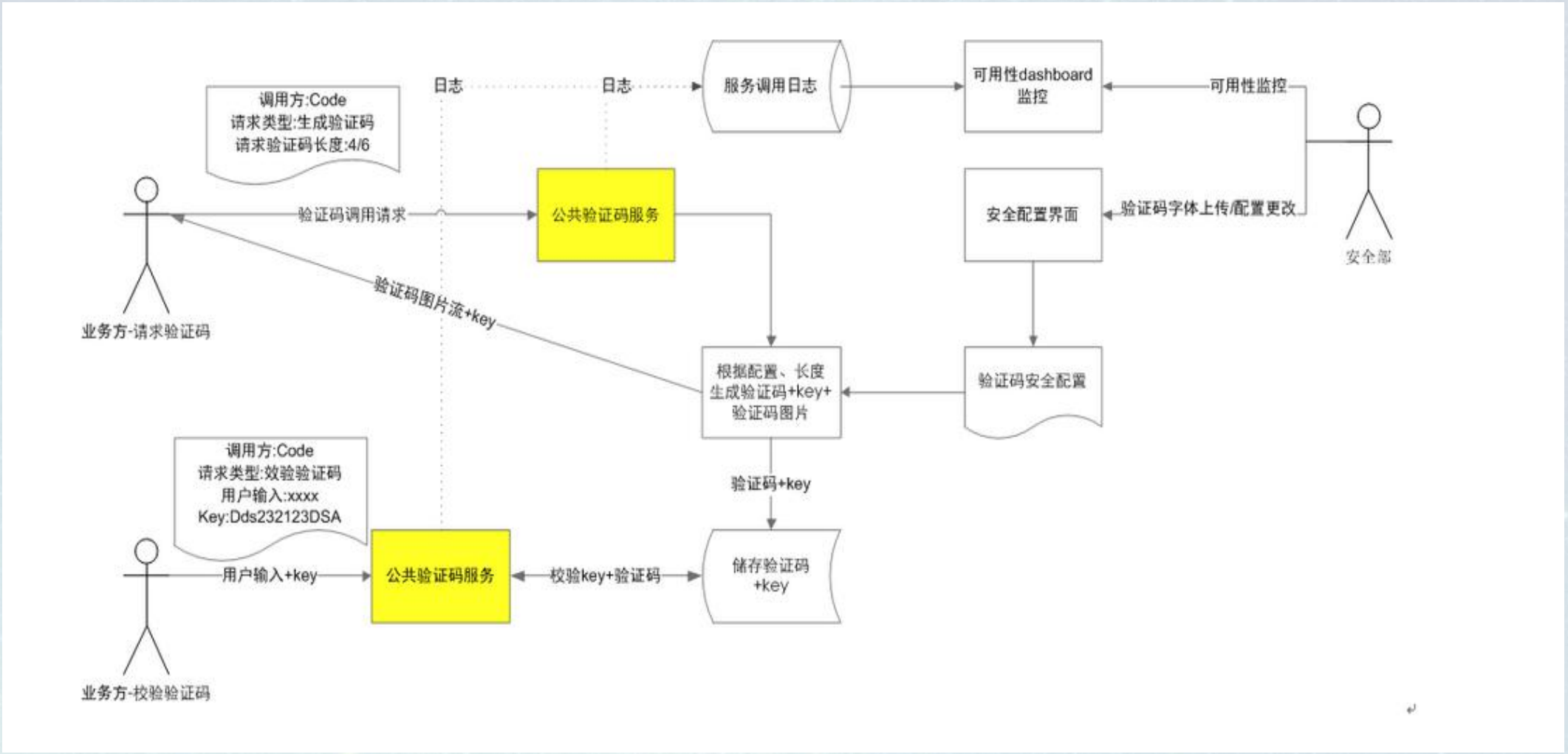
和目前主流互联网OTA企业相比，各种逻辑都偏向于体验和业务，导致了一旦账号被扫之后，后续很难有其他的系统阻止拆信者进行信息采集，资料变更等操作。

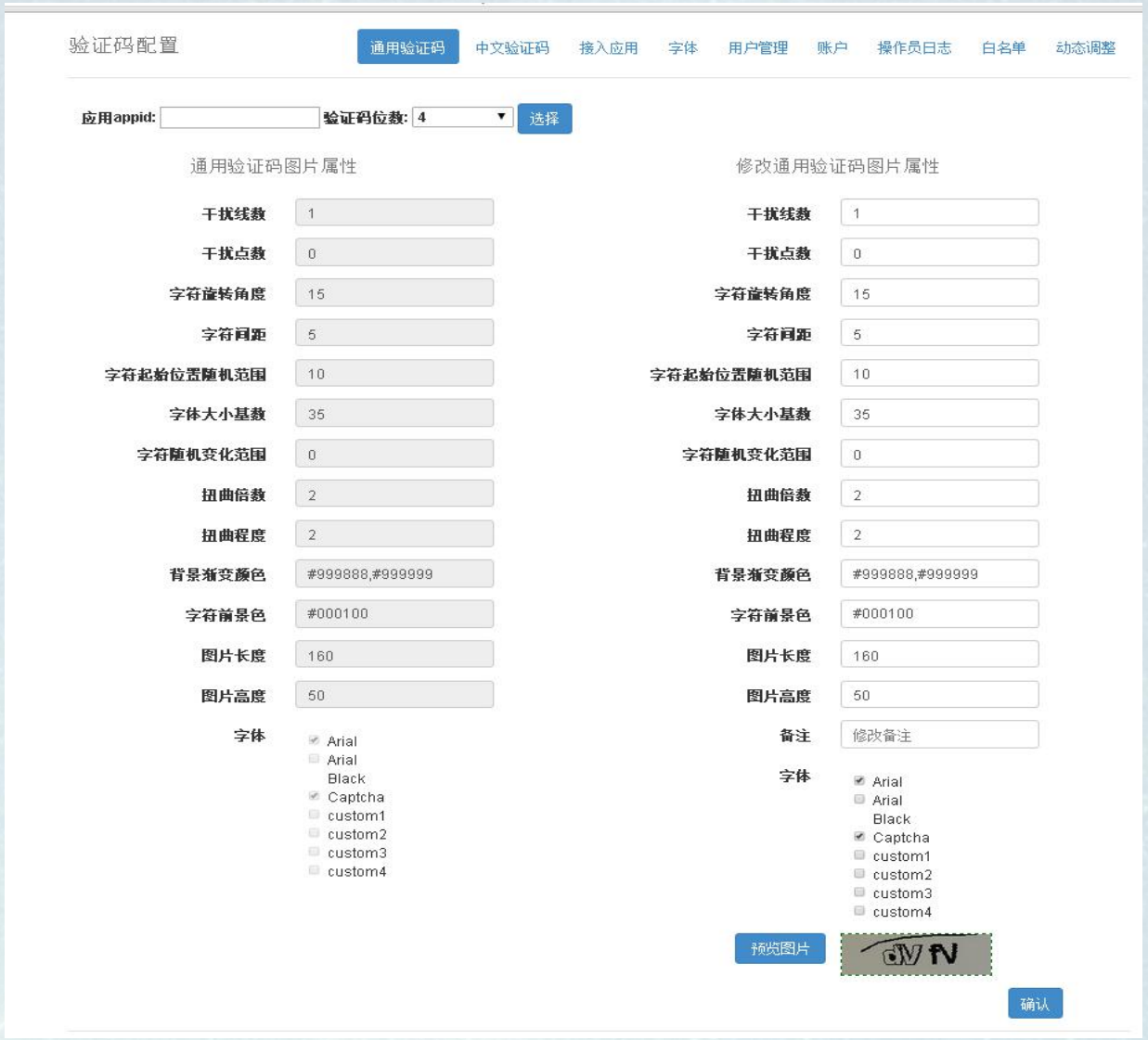
如何去解决

2016携程信息安全沙龙

公共验证码

- 各应用验证码难度解耦
- 自动调节难度和类别
- 各参数独立可变配置
- 服务响应在10ms之内


调用方:Code
请求类型:生成验证码
请求验证码长度:4/6
日志
日志
服务调用日志
可用性dashboard
监控
可用性监控
安全部
业务方-请求验证码
验证码调用请求
公共验证码服务
安全配置界面
验证码字体上传/配置更改
验证码图片流+key
根据配置、长度
生成验证码+key+
验证码图片
验证码安全配置
调用方:Code
请求类型:效验验证码
用户输入:xxxx
Key:Dds232123DSA
验证码+key
业务方-校验验证码
用户输入+key
公共验证码服务
校验key+验证码
储存验证码
+key


验证码配置
通用验证码
中文验证码
接入应用
字体
用户管理
账户
操作员日志
白名单
动态调整
应用appid:
验证码位数: 4
选择
通用验证码图片属性
干扰线数 1
干扰点数 0
字符旋转角度 15
字符间距 5
字符起始位置随机范围 10
字体大小基数 35
字符随机变化范围 0
扭曲倍数 2
扭曲程度 2
背景渐变颜色 #999888,#999999
字符前景色 #000100
图片长度 160
图片高度 50
字体
Arial
Arial Black
Captcha
custom1
custom2
custom3
custom4
修改通用验证码图片属性
干扰线数 1
干扰点数 0
字符旋转角度 15
字符间距 5
字符起始位置随机范围 10
字体大小基数 35
字符随机变化范围 0
扭曲倍数 2
扭曲程度 2
背景渐变颜色 #999888,#999999
字符前景色 #000100
图片长度 160
图片高度 50
备注 修改备注
字体
Arial
Arial Black
Captcha
custom1
custom2
custom3
custom4
预览图片
确认

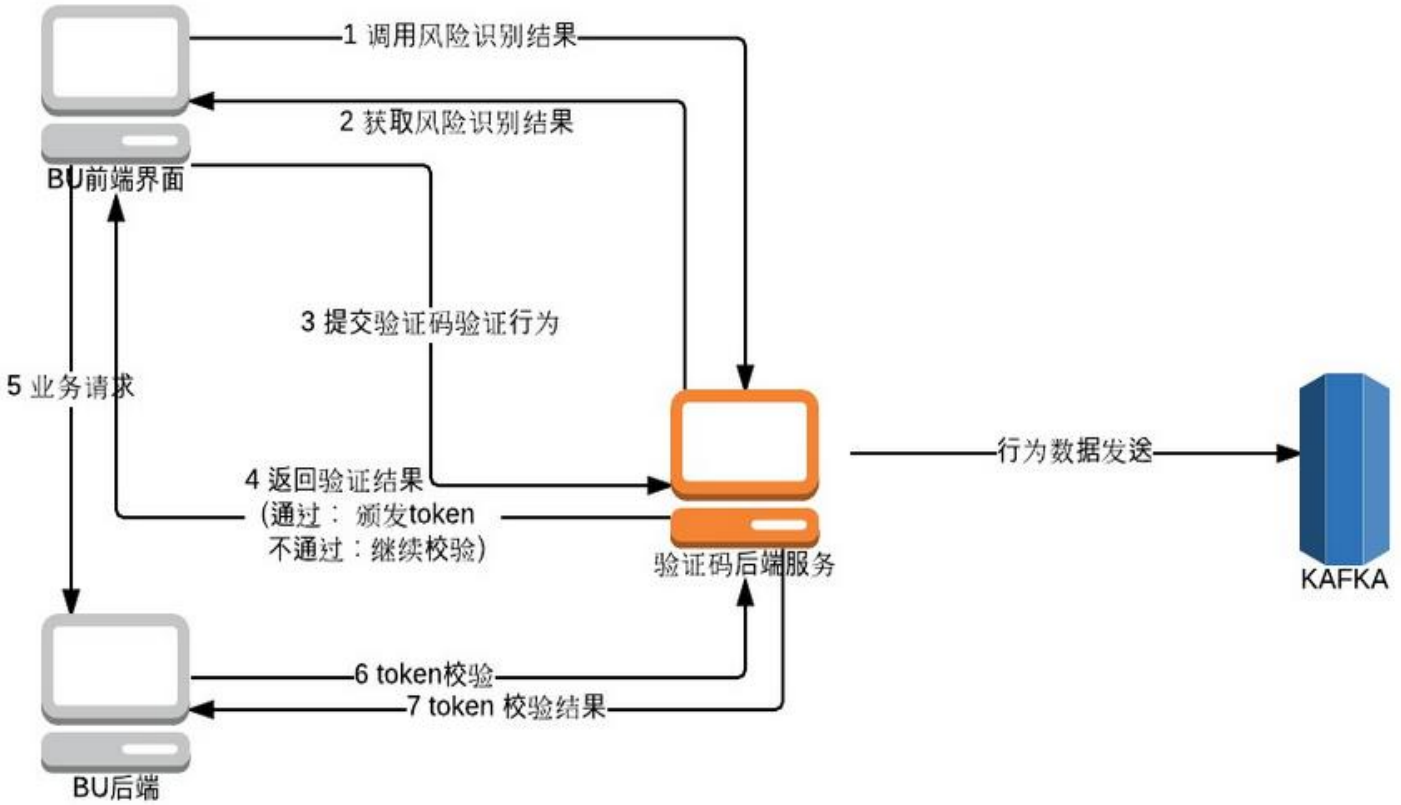
系统概况

- 应对场景：扫号，薅羊毛，信息欺诈
- 效果：英文数字基本可绕过，中文可靠性好
- 成本：每天有50w秒用于输入验证码
- 成功率：基本成功率85%，加权值为90%

问题

- 类别单一，易于破解
- 体验较差，输入验证码时间长
- 识别率存在上限

基础架构


1 调用风险识别结果
2 获取风险识别结果
BU前端界面
3 提交验证码验证行为
5 业务请求
4 返回验证结果
(通过： 颁发token
不通过：继续校验)
验证码后端服务
行为数据发送
KAFKA
6 token校验
7 token 校验结果
BU后端

系统优势

体验由输入变为点击，极大的缩短了验证耗时

支持多级别验证，后台风控多次计算风险，有效阻止异常

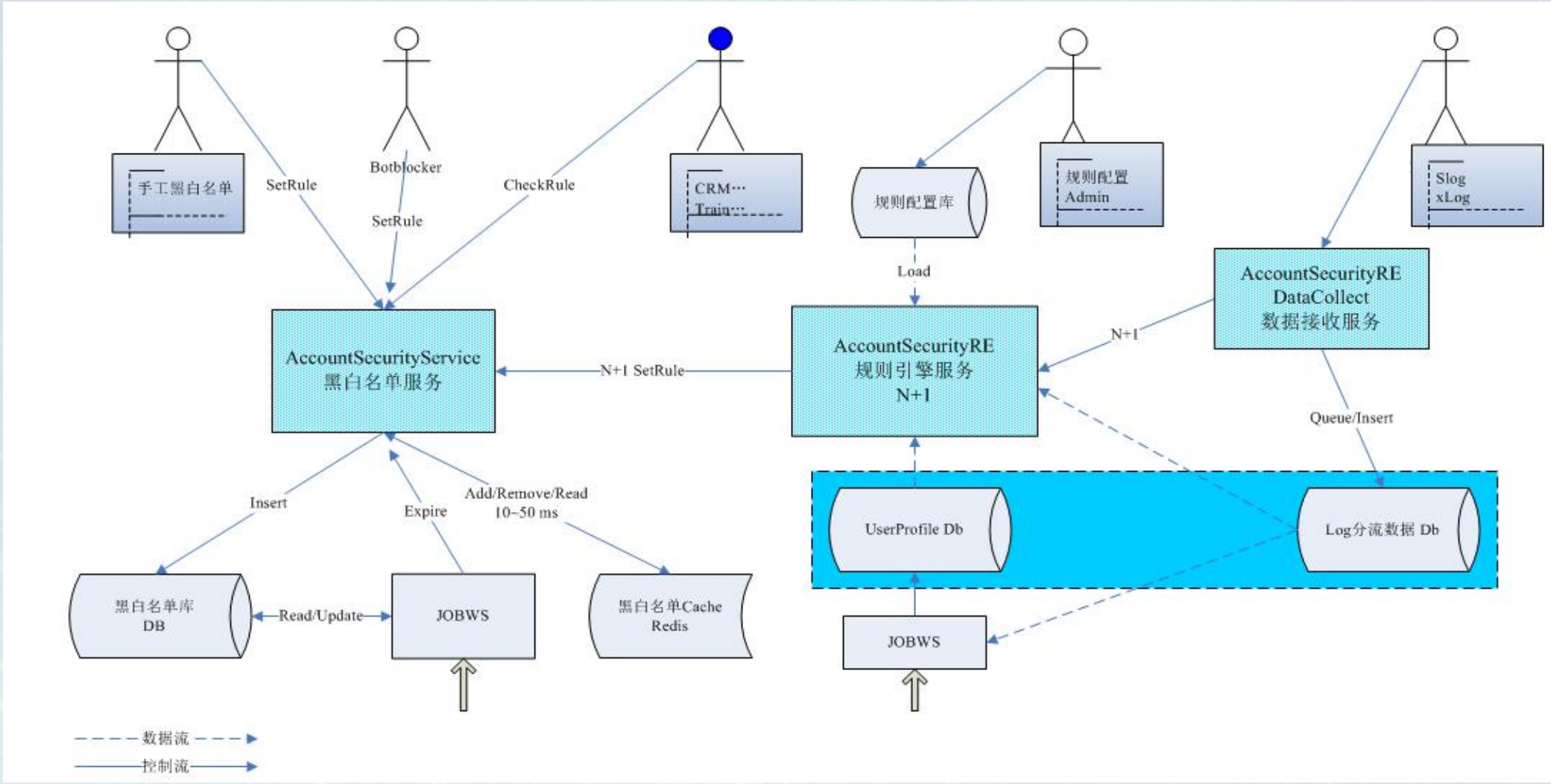
滑块验证加选字验证对真人的识别率理论上可以突破95%

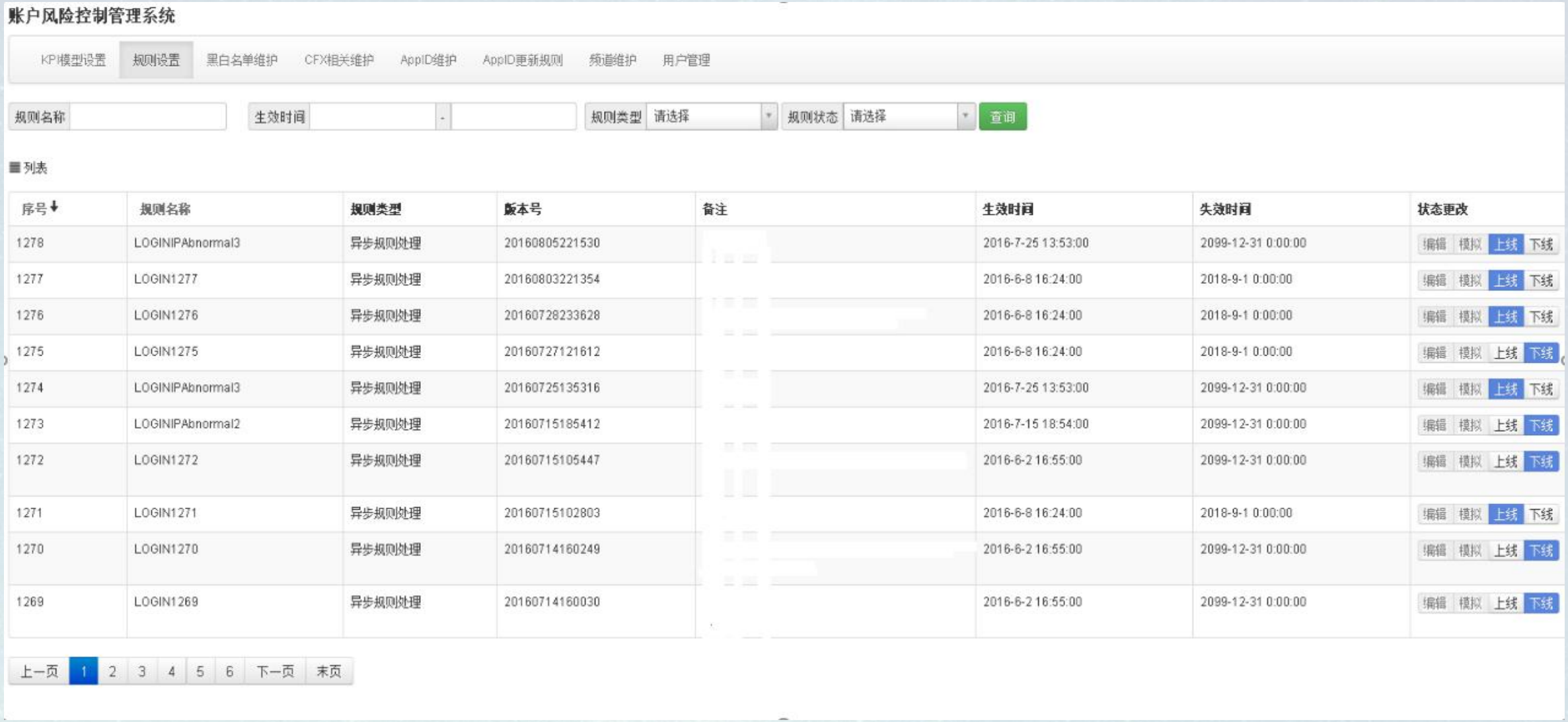
极大的提高了移动端的体验系数，保证了用户体验

在某些无法使用中文的场景，图片选择支持用其他图案

账户风控管理系统

- 实时配置规则
- 异步响应
- A/B testing


手工黑白名单
SetRule
Botblocker
SetRule
CheckRule
CRM…
Train…
规则配置库
规则配置
Admin
Slog
xLog
Load
AccountSecurityRE
DataCollect
数据接收服务
AccountSecurityService
黑白名单服务
N+1 SetRule
AccountSecurityRE
规则引擎服务
N+1
N+1
Queue/Insert
Insert
Expire
Add/Remove/Read
10-50 ms
UserProfile Db
Log分流数据 Db
黑白名单库
DB
Read/Update
JOBWS
黑白名单Cache
Redis
JOBWS
数据流
控制流



账户风险控制管理系统

KPI模型设置 | 规则设置 | 黑白名单维护 | CFX相关维护 | AppID维护 | AppID更新规则 | 频道维护 | 用户管理

规则名称 | 生效时间 - | 规则类型 请选择 | 规则状态 请选择 | 查询

列表

序号	规则名称	规则类型	版本号	备注	生效时间	失效时间	状态更改
1278	LOGINIPAbnormal3	异步规则处理	20160805221530		2016-7-25 13:53:00	2099-12-31 0:00:00	编辑 模拟 上线 下线
1277	LOGIN1277	异步规则处理	20160803221354		2016-6-8 16:24:00	2018-9-1 0:00:00	编辑 模拟 上线 下线
1276	LOGIN1276	异步规则处理	20160728233628		2016-6-8 16:24:00	2018-9-1 0:00:00	编辑 模拟 上线 下线
1275	LOGIN1275	异步规则处理	20160727121612		2016-6-8 16:24:00	2018-9-1 0:00:00	编辑 模拟 上线 下线
1274	LOGINIPAbnormal3	异步规则处理	20160725135316		2016-7-25 13:53:00	2099-12-31 0:00:00	编辑 模拟 上线 下线
1273	LOGINIPAbnormal2	异步规则处理	20160715185412		2016-7-15 18:54:00	2099-12-31 0:00:00	编辑 模拟 上线 下线
1272	LOGIN1272	异步规则处理	20160715105447		2016-6-2 16:55:00	2099-12-31 0:00:00	编辑 模拟 上线 下线
1271	LOGIN1271	异步规则处理	20160715102803		2016-6-8 16:24:00	2018-9-1 0:00:00	编辑 模拟 上线 下线
1270	LOGIN1270	异步规则处理	20160714160249		2016-6-2 16:55:00	2099-12-31 0:00:00	编辑 模拟 上线 下线
1269	LOGIN1269	异步规则处理	20160714160030		2016-6-2 16:55:00	2099-12-31 0:00:00	编辑 模拟 上线 下线

上一页 1 2 3 4 5 6 下一页 末页

系统概况

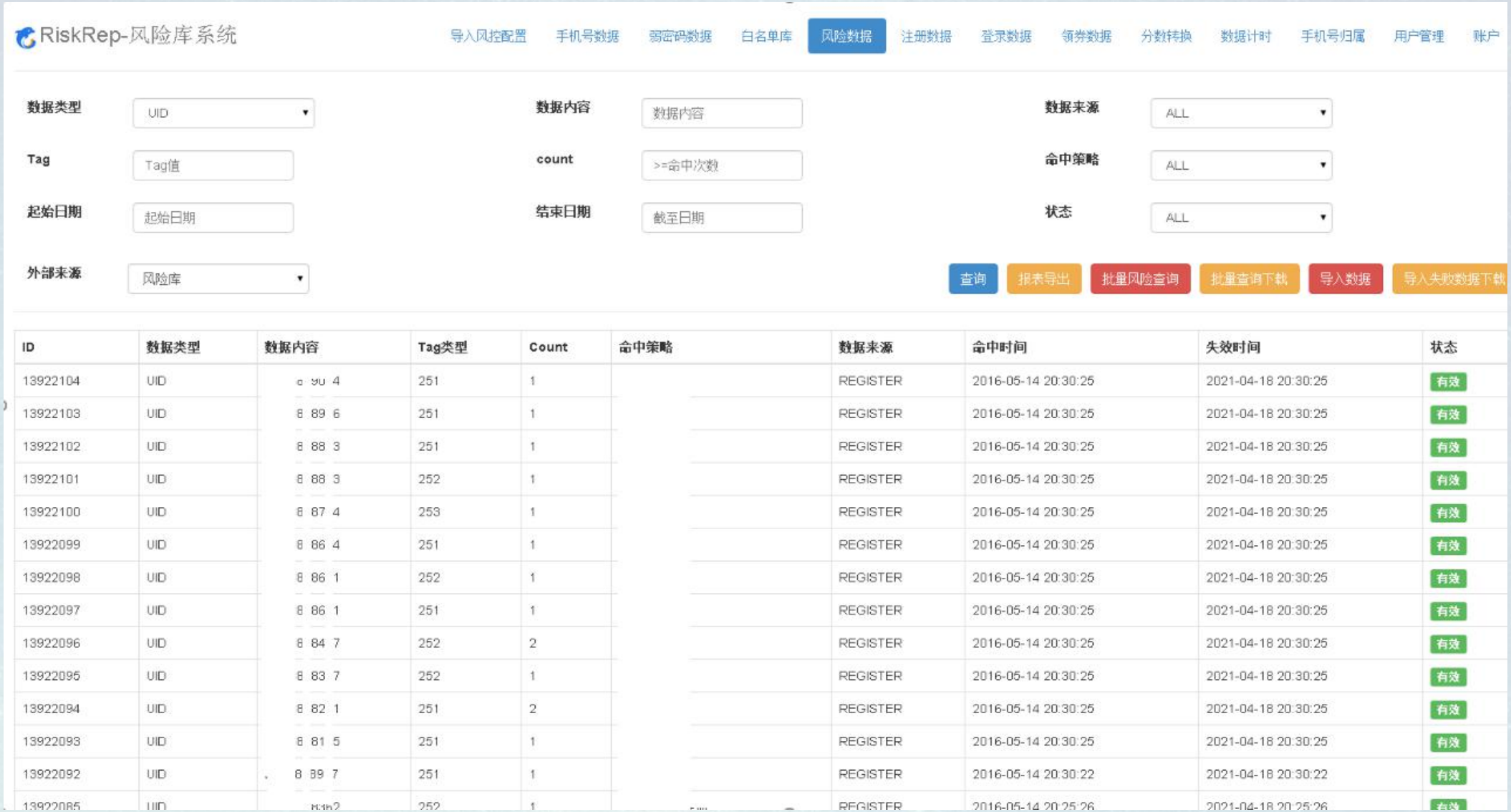
- 应对场景：扫号，薅羊毛
- 效果：日均响应请求 1000w+次，命中规则占比45%
- 性能：平均响应耗时5ms

问题

- 非实时响应
- 非多参数响应
- 无法支持多数据源
- 规则引擎服务写死

黑产数据平台

- 基于离线规则运算
- 数据迭代，分钟级别计算结果
- 与外部黑产数据结合



RiskRep-风险库系统

导入风控配置 | 手机号数据 | 弱密码数据 | 白名单库 | 风险数据 | 注册数据 | 登录数据 | 领券数据 | 分数转换 | 数据计时 | 手机号归属 | 用户管理 | 账户

数据类型: UID　　数据内容: 数据内容　　数据来源: ALL

Tag: Tag值　　count: >=命中次数　　命中策略: ALL

起始日期: 起始日期　　结束日期: 截至日期　　状态: ALL

外部来源: 风险库

查询 | 报表导出 | 批量风险查询 | 批量查询下载 | 导入数据 | 导入失败数据下载

ID	数据类型	数据内容	Tag类型	Count	命中策略	数据来源	命中时间	失效时间	状态
13922104	UID	[illegible]	251	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922103	UID	[illegible]	251	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922102	UID	[illegible]	251	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922101	UID	[illegible]	252	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922100	UID	[illegible]	253	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922099	UID	[illegible]	251	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922098	UID	[illegible]	252	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922097	UID	[illegible]	251	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922096	UID	[illegible]	252	2		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922095	UID	[illegible]	252	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922094	UID	[illegible]	251	2		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922093	UID	[illegible]	251	1		REGISTER	2016-05-14 20:30:25	2021-04-18 20:30:25	有效
13922092	UID	[illegible]	251	1		REGISTER	2016-05-14 20:30:22	2021-04-18 20:30:22	有效
13922085	UID	[illegible]	252	1		REGISTER	2016-05-14 20:25:26	2021-04-18 20:25:26	有效

系统概况

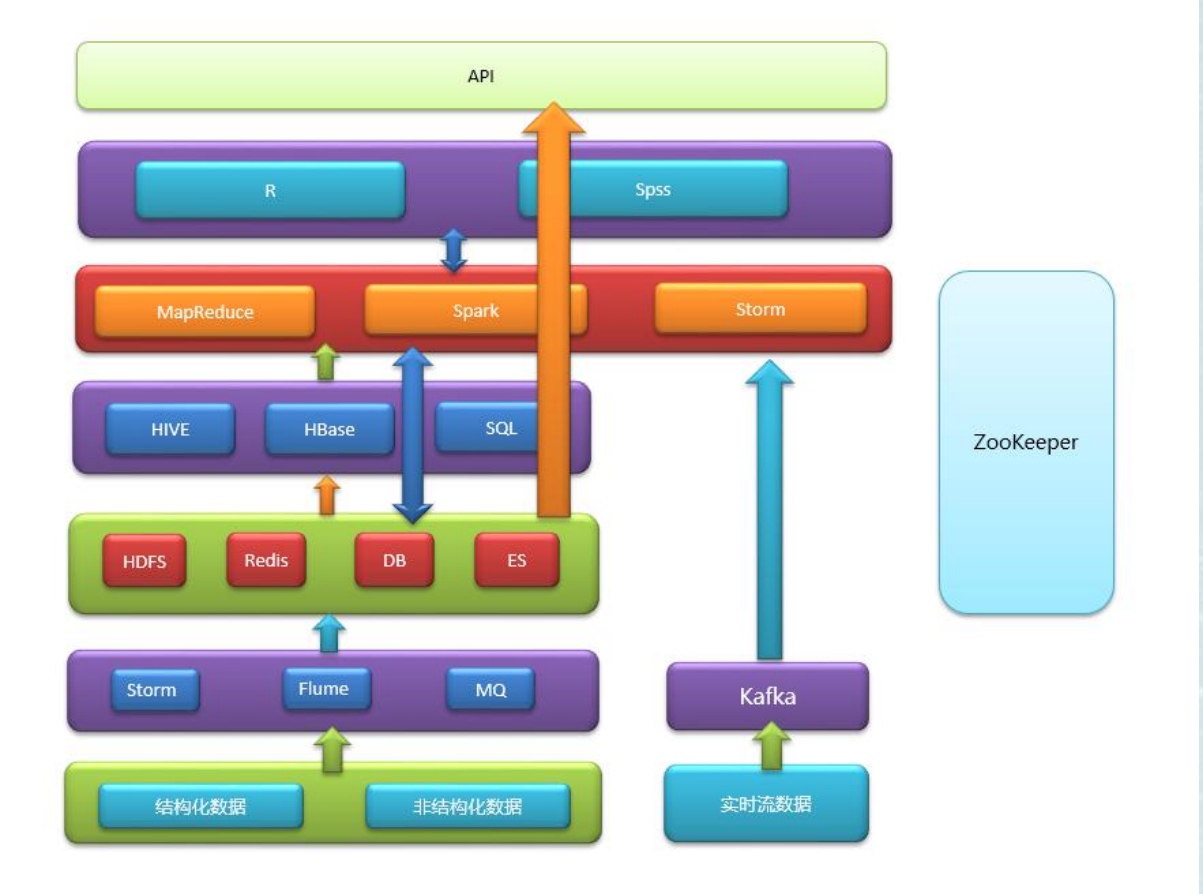
- 应对场景：垃圾注册，扫号，薅羊毛，爬虫
- 对扫号近实时拦截，目前成功账号/IP已经达到了0.7：1
- 反爬主要提供对恶意爬虫的检测，经过A/B测试日均拦截爬虫行为10w次
- 薅羊毛和异常注册识别月均识别超过10w次。

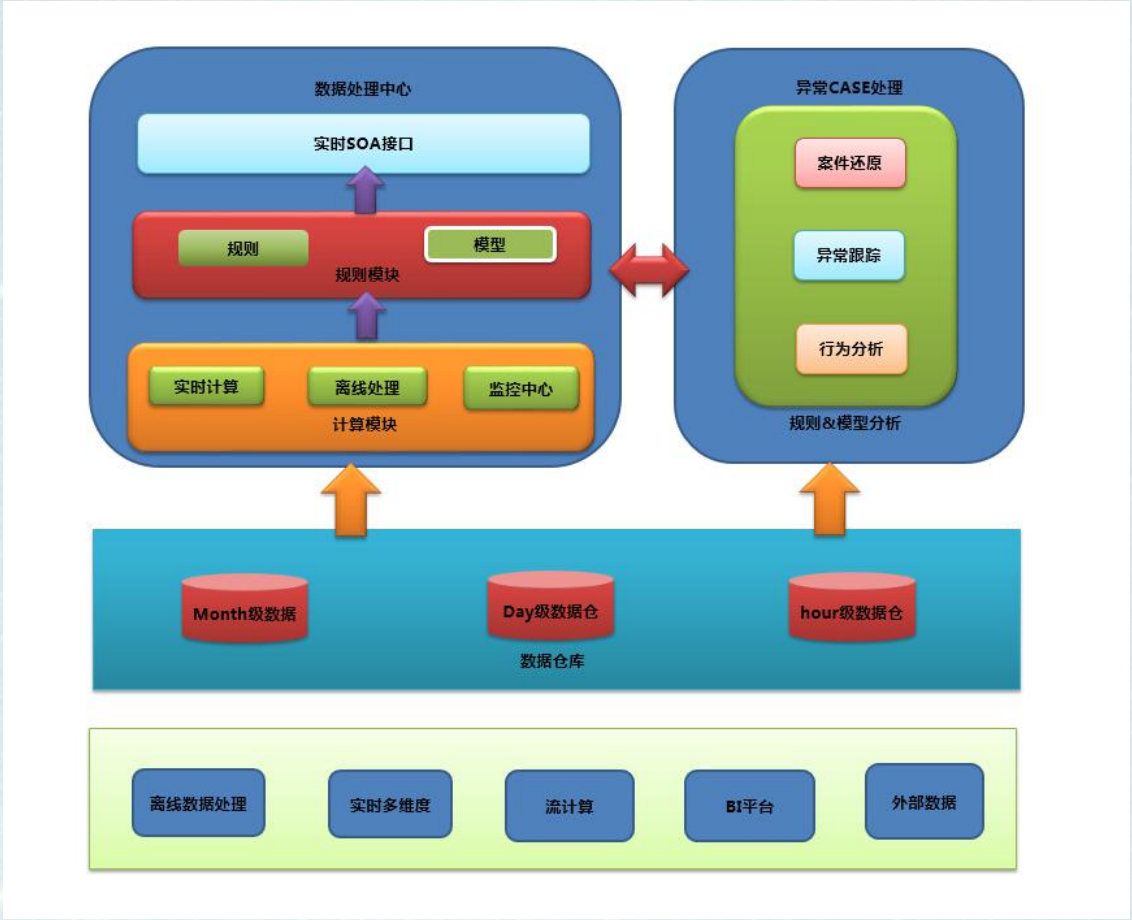
案例：7月27日早晨10点17分，出现大量异常登录请求，黑产平台在10点25分发现异常，迅速介入，10点45将异常登录请求降低到十位数，在11点降低到个位数，对方此时发现无法绕过，直接放弃本次扫号。

本次事件持续40分钟，平台发现异常并介入使用了8分钟，随后20分钟进行了中文验证码干预，效果比较明显，且无需人工进行干预，实现了快速全自动化响应，让扫号无所遁形。

问题

- Sql+DB
- 数据量瓶颈
- 运算效率


API
R
Spss
MapReduce
Spark
Storm
HIVE
HBase
SQL
HDFS
Redis
DB
ES
Storm
Flume
MQ
结构化数据
非结构化数据
Kafka
实时流数据
ZooKeeper


数据处理中心
实时SOA接口
规则
模型
规则模块
实时计算
离线处理
监控中心
计算模块
异常CASE处理
案件还原
异常跟踪
行为分析
规则&模型分析
Month级数据
Day级数据仓
hour级数据仓
数据仓库
离线数据处理
实时多维度
流计算
BI平台
外部数据

系统优势

支持storm流式和spark离线同时给出业务请求的综合安全评分

数据量放大，分析覆盖率增长，可以覆盖更多低频异常操作

规则和模型的同时应用，数据形成互补

标准化体系，无需为了特殊需求反复更新系统架构

未来着力点

2016携程信息安全沙龙

- 移动端的业务安全
- 黑产数据的行业交流
- 重视安全，也要重视体验
- 让用户感知安全，信赖企业


THANKS
Q&A