移动数字金融与电子商务反欺诈

白皮书

（2019年）

牵头编写单位

中国信息通信研究院泰尔终端实验室

中移信息技术有限公司

参与单位

北京数美时代科技有限公司

中国电信世纪龙有限公司

统一推送联盟

电话邦

浙江每日互动网络科技股份有限公司

深圳市和讯华谷信息技术有限公司

泰尔卓信科技（北京）有限公司

北京数字联盟网络科技有限公司

威胁猎人

联洋国融（北京）科技有限公司

北京邮电大学

四川享宇金信金融科技有限公司

目录

图目录

表目录

一、移动数字金融与电子商务中的欺诈现状

随着移动互联网与传统金融和电子商务的深入结合，其不仅带给用户更便捷的使用体验，同时极大地推动了我国数字经济的发展。目前我国在移动互联网服务的发展深度和市场规模都已经领先全球，但与此同时，新的欺诈手段也不断衍生。这种情况不仅给我国广大消费者造成了巨大的经济损失，同时也影响了行业的整体形象，给我国移动互联网的长期健康发展和产业创新带来了诸多消极影响。

按照欺诈对象的不同，欺诈行为主要分为两大类：针对用户的欺诈和针对企业的欺诈。本白皮书主要研究针对企业的欺诈行为及其防范方法。

1.1 移动数字金融与电子商务欺诈概述

金融和商品交易是现代经济体系的核心。随着信息技术的发展，金融和商品交易也在逐步信息化，形成了数字金融和电子商务的模式。无论是服务方式、获客渠道都基于现有的信息化基础设施，极大提升了传统经济活动的服务效率，降低了服务成本。然而，与此同时信息服务也给欺诈行为带来了更多的手段和渠道，使得传统欺诈行为的危害大大提升。

以网贷平台为例，截止 2018 年末，累计出现问题的平台数量超过 4000 家，占网贷平台总量的 70%以上。而在电子商务领域，根据

Pymnts.com 在 2017 年 10 月发布的一份全球电子商务欺诈报告，电商欺诈将导致全球电商市场在 2017 年损失 580 亿美元。在此背后的“黑色产业”肆虐发展，已经渗透到账号注册、身份伪造、宣传导流、借贷支付等各个环节。据估计，相关“黑色产业”从业人员超过 500 万，涉及金额达到千亿级别。总体而言，数字金融和电子商务是欺诈行为高发的“重灾区”，成为形形色色的黑色产业攻击的主要对象。

整个市场流量“移动化”的背景下，不论是传统线下业务还是原本由 PC 互联网承载的业务，都在逐步向移动端拓展。而其在整个移动互联网业务中，数字金融和电子商务是两个非常重要的领域，与广大居民的日常生活息息相关。与此同时，上述两个领域所暴露的欺诈风险也越来越严峻。

具体而言当前的移动欺诈主要包括以下几种形式：

1.1.1 营销活动欺诈

营销活动欺诈指，在企业进行新用户获客及老用户唤醒时所采取的如红包、优惠券等运营成本，被黑灰产利用技术手段不正当获利，导致营销活动失败的场景。

在营销活动欺诈中，存在羊毛党和黄牛党两种关键角色。

图 1 营销活动反欺诈示例

- 羊毛党：操纵大量账号仿冒新用户，参与营销活动，获取优惠券奖励。或者通过收取费用代人下单，从而获取利益。

- 黄牛党：操纵大量账号参与营销活动，活动购买资格，购买后，高价卖给其他用户，从而获取利益。对于比较稀缺的、价值比较高的商品，会出现黄牛党。

1.1.2 渠道流量欺诈

渠道流量欺诈指，黑灰产利用技术手段仿冒移动应用新增用户，独自或与第三方推广平台合作，共同骗取移动互联网应用（App）市场运营成本的场景。

据数美科技统计，2017 年全球范围内 App 安装欺诈占总 App 推广安装量的 7.8%左右，亚洲地区 App 安装欺诈占同地区总 App 推广安装量的 11%～12%左右。保守估计 2017 年全球由于渠道流量作弊导致的损失高达 11～13 亿美元。

目前，随着移动互联网的高速发展，渠道流量作弊也呈现出快速

增长的趋势。

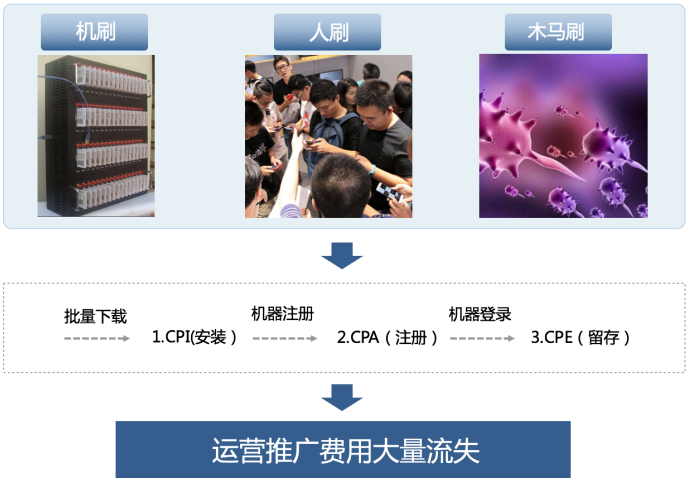

图 2 渠道流量反欺诈示例

如上图所示，App 安装渠道流量作弊有不同的形式，其中常见的几种：

- 机刷：通过批量地虚拟机、篡改设备等手段，刷安装激活；
- 人刷：通过做奖励任务形式，人肉刷安装激活；
- 木马刷：通过感染移动设备，在正常手机后台偷偷刷下载激活；
- 点击劫持：通过恶意软件，当检测到用户下载安装某 App 时，发出点击记录。

除了这些手段，App 安装渠道流量作弊也越来越隐蔽，使得检测难度越来越大，常见的伪装包括：

- 通过代理 IP、位置模拟、设备型号伪装等，让群控设备看起来像是自然分布；
- 在安装激活后，继续模拟后续的 App 内用户行为，使得留存

率看起来正常；

1.1.3 虚假用户裂变欺诈

虚假用户裂变欺诈是指 App 采用“用户裂变”的方式来进行推广获客时，黑产通过控制大量假账号，骗取平台拉新补贴的场景。

当前，移动互联网用户流量红利渐渐消失，App 常常使用“用户裂变”的方式进行拉新获客，以期降低流量获取的费用。如下图所示，所谓用户裂变，就是将 App 已有用户都是获取新用户的渠道，即通过一定的激励措施，刺激已有用户通过自己的关系链帮助平台拉新。

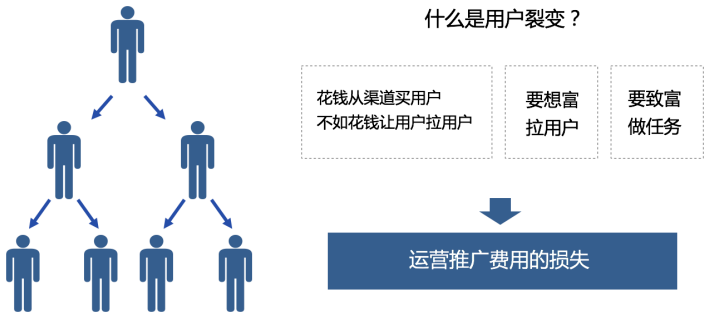

图 3 虚假用户裂变反欺诈示例

该场景中，常依靠邀请码等方式建立师徒关系后，要求徒弟或者师傅做任务，才能获得所有奖励或者获得提现资格。例如，拉取新用户以后，需要徒弟每天完成一定的任务量才可以获得奖励，一般需坚持 7 天时间才能将拉新奖励的 3-8 元拿完。黑产通常通过注册大量的假账号，骗取平台拉新补贴。

1.1.4 盗取信息欺诈

盗取信息欺诈指，欺诈团伙通过高额利息、高价值奖品、高额度折扣等虚假宣传欺骗网络用户，并要求用户填写个人信息，从而实现非法盗取用户信息的目的。

相关欺诈作案手段多样，难以发现。而非法获取的公民信息又常常通过非法转卖的方式流入地下黑产，给居民造成巨大的隐患。

1.1.5 恶意交易欺诈

恶意交易欺诈指，黑灰产利用移动互联网交易的便利性，在交易中的货到付款、退、换货政策等环节中，利用漏洞进行牟利。

不同于传统的线下交易模式，在移动互联网的线上交易中，由于交易实现的便利性，交易生成的过程得到了大大简化。这在给消费带来便利的同时，也使得以欺诈为目的的大规模恶意交易成为可能。这些交易通常利用货到付款、退、换货政策等电子商务交易机制中的漏洞进行牟利，或以让商家受到损失为目的。这类欺诈通常隐蔽性更高，其中很大部分并不直接以牟利为主要目标，而来源于针对于对商业同行的恶意攻击。这类有组织的恶意交易近年来增长迅速，相关欺诈的恶意蔓延，可能对我国移动互联网健康的商业环境造成长期重大的消极影响。

1.1.6 金融支付欺诈

金融支付欺诈指，利用不正当的技术手段在支付的各个环节谋取不正当利益的行为。

目前，移动金融尤其是移动支付已经成为居民日常生活中不可或缺的组成部分。目前，基于金融和支付的欺诈也在日益增长。包括利用的支付系统的漏洞在用户不知情的情况下非法盗取用户资金；通过伪造网站、公司、项目等手段骗取用户资金；通过一些第三方支付平台发行的商户的 POS 机虚构交易套现；将非法所得的资金转移到第三方支付平台账户，在线购买游戏点卡、比特币、手机充值卡等物品，再对外销售进行洗钱等活动。这些行为严重扰乱了金融和社会秩序。

1.1.7 网络刷单欺诈

网络刷单欺诈指，灰产模拟活跃用户对商品评论或购买数量进行恶意操纵，从而导致消费者受到欺骗或商家受到损失。

随着移动互联网业务日益发达，评论和反馈机制对于商品质量和服务提升起到了重要作用，用户评论和购买数量等数据已经成为用户做出选择的重要依据。正因如此，相关业务也伴随着重大的经济价值，成为另一个欺诈高发的领域。一些商家有意的恶意操纵评论，误导消费者做出错误选择，严重破坏了整个移动互联网商业生态的信用体系。目前，刷好评、炒信用已经衍生为灰色产业链，各种刷单、刷信誉等兼职层出不穷。在这种情况下，很容易产业 “劣币驱逐良币”的现

象，某些卖家刷好评、刷信誉度的同时，遵守规则的商家利益就会受到侵害，从而对我国数字经济的长期健康发展产生造成不良影响。

网络刷单团伙的特点：操纵大量的账号，并通过运营刷手群或直接利用软件工具来实现对平台玩家的粉丝数/评论数等多项指标进行刷榜造假；与有需求用户交易，从而谋取利益。

图 4 网络刷单欺诈示例

1.1.8 电信欺诈

电信诈骗主要通过电话、短信以及互联网联系作为主要手段的诈骗案件，意在获取被害人的财产、银行账户等隐私信息。常见的手法有：①冒充熟人进行诈骗：如，冒充公司领导、摸清公司人员架构后向财务人员发送转账汇款指令；②以中奖、退税、积分兑换等馅饼类为由，进行诈骗：如，事先获得事主购买的房产、汽车等信息，以税收政策调整办理退税为由，诱骗事主转账到指定账户；③冒充公检法、公安局等政府机构，进行诈骗：如，通过收集的受害者的隐私信息如身份证号、工作单位、住址等，获得初步信任，再通过改号软件伪装

为警方电话，假称受害者涉嫌洗钱、非法集资等重大犯罪案件，诱导其一步步将资金转入指定账户。

1.1.9 网贷欺诈

网贷欺诈风险是指，申请人的还款能力无法通过互联网有效远程判断，申请人利用线上申请环节的漏洞伪造数据故意违约或线上黑色产业链利用技术手段劫持互联网贷款平台信息恶意进行团伙欺诈行为。

随着互联网+模式的深入各个行业，互联网贷款市场也在不断扩大，随之而来的是大规模的线上逾期风险和线上黑色产业野蛮生长。主要欺诈手段有；申领大量手机号码，同时利用这些非常用号码进行大量刷量消费从而提高信用评级；通过技术手段修改伪造身份信息、手机设备信息、位置信息达到骗取贷款并躲避贷后催收的目的；利用公共信用信息更新缓慢的时间差同时申请多家平台贷款，恶意透支信用度。

1.1.10 优质内容爬取欺诈

优质内容爬取欺诈，是指通过网络爬虫（又称网络蜘蛛），按照某种规则在网络上爬取所需内容的脚本程序。

对于被爬取内容的各种资讯类 App 来说，损失非常巨大。这些平台雇佣大量编辑人员，投入大量时间、金钱成本、写出\运营出的高质量内容，却很快被爬虫窃取，形同侵权。

再以各类出行机票类 App 为例，此类 App 上的机票价格大都采用动态定价的方式，服务器会结合当下浏览量判定机票的抢手程度并且调整价格。这时如果有大量爬虫在浏览 App，算法就会给出和实际情况并不符合的定价，这也会损伤消费者购买到廉价产品的权益。

爬虫带来的危害远非如此，爬虫的行为会极大地增加数据分析难度，文章浏览量的失实让我们误判人们对新闻事实的关注程度、爬虫衍生出的虚拟 IP 需要在数据清洗时剔除……技术越高超的爬虫，在行为模式上就越接近真人，也就更加增加数据分析时的难度。久而久之，那些我们以为从人类行为中寻找规律的算法，反而寻找到的是机器人的行为规律。

总体而言，爬虫盗取内容和数据的行为对企业危害甚大，不仅会降低企业内容新鲜度，甚至侵犯企业敏感数据、增加企业运营风险。

1.2 移动数字金融和电子商务领域的反欺诈场景

1.2.1 移动用户的身份判断

现在绝大部份 APP 和网站在注册时都是需要利用手机号、IP 等基础资源。大部分欺诈行为也是首先囤积虚假账号然后进行后续的针对不同场景的欺诈行为。特别是当电商行业有某个重要的促销活动前（例如天猫双十一、京东 618），黑灰产会进行大量囤积账号的行为；其中去年双十一时，电商平台遭遇的虚假注册账号已达到 160 万次。虚假账号的识别是反欺诈场景的基础，也是企业对抗黑灰产的基础。

1.2.2 移动欺诈的状况评估

在移动欺诈的场景中，企业自身对欺诈状况的掌控是至关重要的。反欺诈与传统安全最大的区别在于，传统安全是边界安全，而业务场景下的反欺诈安全是叫安全可控。而反欺诈场景下的安全更多关注的不是企业是否存在容易被攻击的漏洞，而是企业的业务逻辑是否容易被黑灰产利用，黑灰产在企业各个场景下的欺诈成本有多少。例如，最开始黑灰产赚取 100 元只用消耗 1 元的成本，在企业上线了很多策略后，黑灰产仍然能投入 1 元赚取 100 元的话，那说明这些策略的堆积并没有发挥作用。因此在移动欺诈场景下的状况评估，是通过对黑灰产攻防成本的监测和企业业务逻辑漏洞及流程缺陷进行监测，了解企业的移动欺诈的状况。

可以从以下几个维度去判断：

- 虚假账号量
- 注册风险
- 登录风险
- 流量欺诈风险
- 内容风险
- 活动风险
- 数据风险
- 设备风险

1.2.3 移动欺诈的行为判断

黑灰产在进行欺诈行为时，都会有一定的规律行为，而为了投入产出比的最大化，往往会利用自动化工具和脚本去运行这些固定的操作行为，让其看起来更像一个正常人的操作，避开企业的风控策略。

例如某刷单厂商的整个刷单流程需要经过待刷物品资料整理，模拟浏览，模拟聊天，付款，确认好评等步骤完成。

从以上整个流程来看，黑灰产进行刷单欺诈时其模仿正常用户的行为非常细致，并且所使用的电商账号大多跟正常买家无异，所以电商平台需要通过多维度特征加行为分析才能够有效识别出刷单的欺诈行为。

除此之外，还可以通过设备维度判定欺诈行为，可借助反欺诈工具，如“可信 ID”，判定移动设备唯一性。若设备是真实的，其背后的用户可能是真实的；若设备是虚假的，其背后的用户则存在一定的风险。

1.3 数字欺诈对我国经济的影响与分析

1.3.1 当前网络欺诈的现状

目前，我国互联网在用户规模、业务模式创新、新零售与文娱产业等多个方面持续保持着快速增长，根据 CNNIC 发布的《第四十一次中国互联网络发展状况统计报告》，中国网民规模在已经超过 8 亿，普及率超过 60%，其中移动互联网用户占比达超过 90%。与此同时，

基于移动互联网的应用场景不断丰富，移动支付比例已经超过70%，网络娱乐用户规模持续高速增长，网络直播用户年增长率超过100%，电子商务、网络游戏、网络广告收入水平增速均在20%以上。

但在享受互联网带来的生活品质提高的同时，我国用户也在遭遇多种有组织的网络欺诈行为。根据CNNIC数据统计，在2018年30%以上的网民遭遇了个人信息泄漏，超过25%的用户遭遇网上诈骗，23.8%的用户遭遇了病毒或木马攻击，19.2%的用户账号或密码被盗。全年境内感染病的移动毒终端累计超过3000万台，国内被篡改网站累计超过7万个，安全漏洞累计18901个，其中高危系统漏洞累计7654个，较2017年增长了31%。

对互联网行业的企业而言，灰色产业带来的不仅对正常业务的打扰更是真实的经济损失。对于很多移动互联网企业，在开拓市场之初，纷纷推出了花样繁多的优惠活动。这些优惠在吸引了众多用户改变使用习惯的同时，也成为了灰色产业从业人员的目标。大量不法分子利用作弊软件与作弊硬件，通过绕过监控规则，通过虚假身份欺诈套利，获利丰厚，甚至渐渐形成了整套的灰色产业链。而这些产业因为自身的隐藏性与反侦察性并不为社会公众所感知。

1.3.2 移动互联网欺诈的模型和结果分析

数字金融及电子商务中的欺诈行为中，不管是企业还是消费者都可能成为受害方。在进行服务的过程中，当一方采取虚假信息或者其他不正当手段欺骗、误导另一方，使得其在违背其真实意图的情况下

完成业务或服务，就构成了欺诈行为。

广义上来看，欺诈也是一种经济行为，并且通常可以带来超额利润，从而引起更多参与者的加入。从这个角度来看，是一种欺诈也可以看作是一种“反应—扩散”模型。反应扩散模型是一个同时考虑扩散、迁移和增长的模型。具体而言，构成移动互联网服务欺诈的组成部分如图 2 所示：

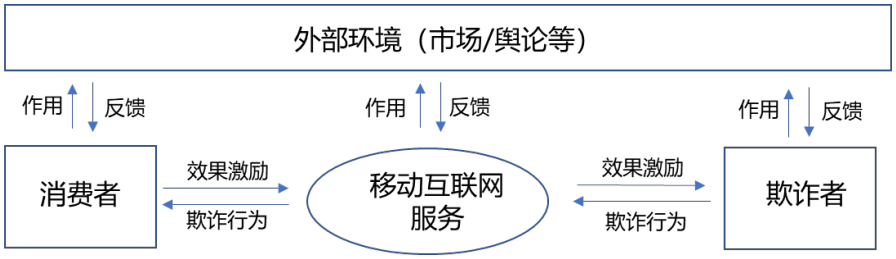

图 5 反欺诈扩散模型示例

在欺诈者与受害者（商家和用户都可能成为欺诈者）的行为模式影响下，通过移动互联网服务为平台，与外部环境相互作用，形成欺诈者与受害者的反应扩散模式，最终使得移动互联网反欺诈行为经过作用与反馈的不断迭代，形成较为成熟的模式。

从这种角度来看，欺诈模式的形成类似于传染病的传播模型，作为欺诈者和受害者都在不断的迁移、发展和退出。同时，需要假设反应扩散的过程使在移动互联网的“状态空间”中连续变化，个体（欺诈行为）随机迁移并以相同的概率向各个方向（移动互联网的不同业务）扩散。其中，反应是指在这一过程中，欺诈者和被欺诈者之间的相互作用——一方面是在欺诈者数量S，欺诈所获得的收益I的条件下，新的欺诈行者被“激发”出来，并表示为。以及由于某种欺诈行

为带来的回报而诱使其他欺诈人员参与，并使得欺诈“竞争激烈”并引发更严格的监管，使得欺诈代价提升$G(S,I)$。

作为一个典型的反应扩散模型，可以看到欺诈者数量S，欺诈所获得的收益I是相互耦合的，欺诈者的数量会影响欺诈的收益，反过来欺诈的收益情况又会“激发”新的欺诈者。与此同时，和任何的经济行为一样，无论是欺诈者数量S还是欺诈所获得的收益I，都会随着时间不断“扩散”——在不同领域中，欺诈者和收益都将沿着“梯度”方向进行扩散，向着收益更大的方向上演化。

在均匀的初始条件下，经过长时间的演化，最终通常会呈现出某些宏观的空间特性（例如在电子商务和数字金融领域，欺诈发生的频率明显增高）。

具体推导过程可参照附录 A。

根据我国电子商务、数字金融的市场规模以及我国近年来上述领域由于欺诈造成的损失统计值来插值计算上述几个系数。根据本白皮书参与单位联合提供的数据，首先我们统计出可疑账号的数量，从 2014 年到 2018 年依次为 330 万、412 万、534 万、723 万、1060 万 。

为了估算出从 2014-2018 年欺诈造成的总体（直接）损失，我们选取了 653 组调研和实际案例，并依此估算欺诈造成的损失。根据各年的案例估算每欺诈账户每年获得的收益为 3.13 万、3.98 万、4.51 万、4.24 万、3.31 万。

表 1 电子商务及欺诈市场明细

时间	电子商务市场规模（万亿）	数字金融市场规模（万亿）	欺诈造成的损失（亿）	疑似欺诈者账号数量（万）	单账户欺诈造成年均的损失（万元）
2014	12.6	8.1	1032	330	3.13
2015	16.9	11.2	1642	412	3.98
2016	19.3	15.7	2411	534	4.51
2017	22.4	16.1	3072	723	4.24
2018	25.2	17.2	3513	1060	3.31

不难看到，随着疑似欺诈账户的增加，使得欺诈的同行“竞争”更加激烈，并使得相关企业投入更多资源用于反欺诈，造成了欺诈的收益率降低，这也从侧面验证了本文所提出的欺诈数学模型的有效性。同时，根据上述数据对本本模型进行参数拟合，可以得到

表 2 拟合参数结果

参数	α	β	p	q
数值	0.002	0.013	0.15	0.31

基于上述参数，如果我国移动互联网产业发展以及欺诈情况继续维持当前现状，则未来可能造成的损失预测如下：

表 3 预测损失结果

年份	疑似欺诈者账号数量（万）	单账户欺诈造成年均的损失（万元）	欺诈造成的总损失（亿）
2019	1250	3.09	3870
2020	1310	3.93	5150
2021	1740	3.41	5940
2022	2300	3.08	7100

从统计结果来看，2018 年我国移动互联网欺诈造成的损失大约相当于我国当年 GDP 的 0.3%。但是，如果 GDP 保持在 6.5%的速度发展，则按上述预测，从 2019 年到 2022 年，互联网欺诈可能造成的损失占 GDP 的比例依次为

表 4 欺诈损失 GDP 占比预测

年份	2019	2020	2021	2022
欺诈造成的损失占 GDP 的比例	0.4%	0.5%	0.55%	0.61%

注意实际情况中，还存在大量未被发现的欺诈账号和欺诈行为，同时随着我国数字经济的发展，新的业务和新的欺诈形式也可能同时出现。因此，在实际中欺诈造成的损失很可能比本文推算的更大。

二、黑产欺诈态势分析

2.1 黑产欺诈问题当前态势

随着“互联网+”的快速发展，包括金融、教育、医疗、零售、出行等在内，越来越多的行业与互联网深度结合，为消费者提供越来越便捷的服务。与此同时，越来越多的黑灰产也盯上了这里的巨大利益，网络欺诈呈现出愈演愈烈的趋势。

研究报告《欺诈经济学：规避快速增长和创新中的风险》中指出，黑产从业人数高达 150 万，2015 年网络欺诈损失占 GDP 比例多达 0.63%，约 4000 多亿人民币。这些只是欺诈造成的直接经济损失，在这之外，欺诈造成的客户信任、品牌形象等方面的损失则难以衡量。

黑产欺诈问题具体包括：支付诈骗交易规模逐年增长，互联网恶意机器流量规模及增长率趋势逐年增长，黑产广告造成的损失逐年增长。

具体而言，白皮书试图从黑产的涵盖范围，黑产的技术手段，黑产的实施方式及黑产的产业链条等四个维度论述黑产的欺诈态势。

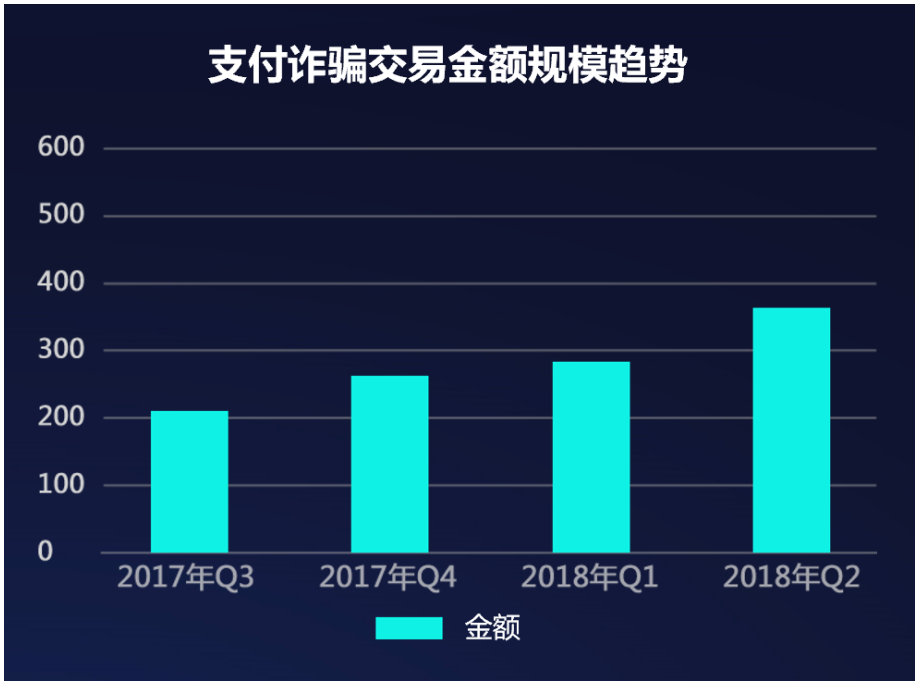

图 6 支付诈骗趋势（中国信息通信研究院）

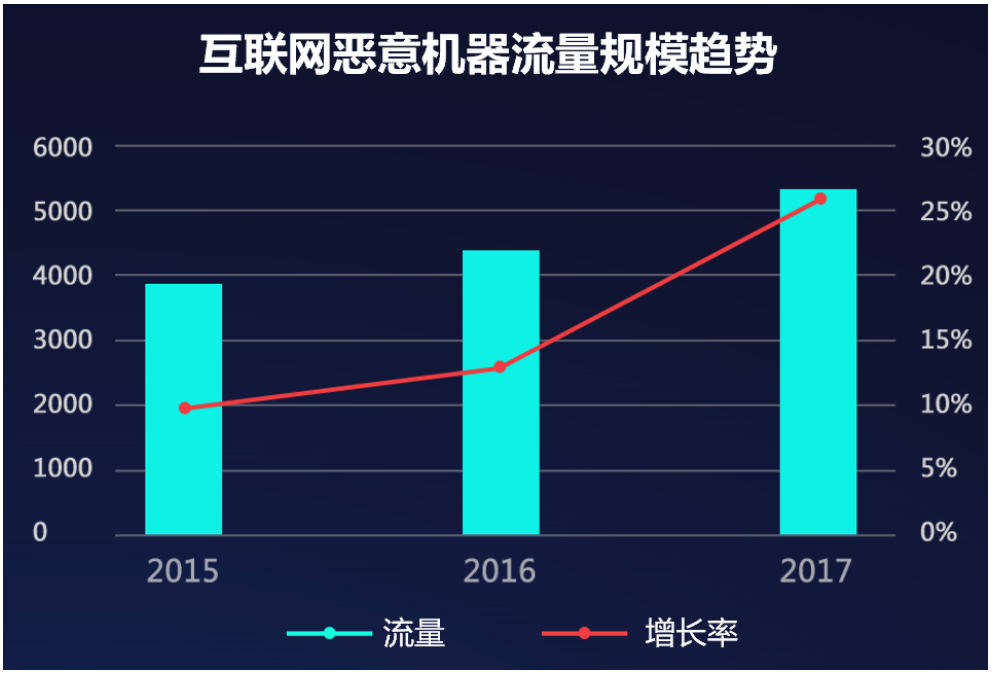

图 7 恶意机器流量趋势（CNNIC）

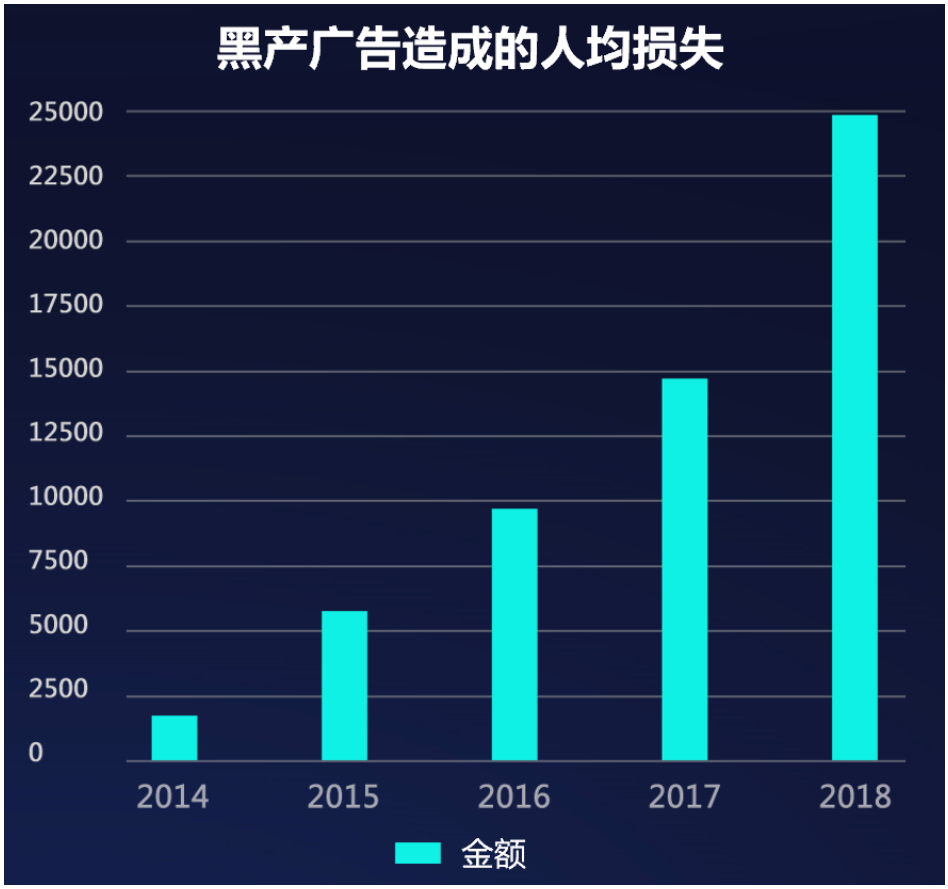

图 8 黑产广告造成的人均损失《2018 年网络诈骗趋势研究报告》

2.1.1 黑产无孔不入，损失巨大

如下图所示，黑产遍布金融，电商/新零售，社交，出行，游戏等多个领域，多个行业，造成的损失高达 4000 亿，破坏极大。在银行转账，反欺诈，信用卡盗刷，刷帮刷单，广告导流，虚假用户裂变，盗刷积分，渠道流量作弊等多场景下，都有黑产的相关身影。

图 9 诈骗场景示例

2.1.2 黑产作恶多端，手段多样

黑产手法多种多样，包括伪基站、猫池、卡池、设备农场、打码平台、积分墙等。

图 10 黑产手法及设备

2.1.3 黑产灵活多变，进化神速

黑产 7*24 小时实时盯守，发现漏洞及时行动，发现被拦截后及时更改策略，进化神速；同时黑产并且遍布全球，全球协同进化，技术全球范围转播，升级速度快。

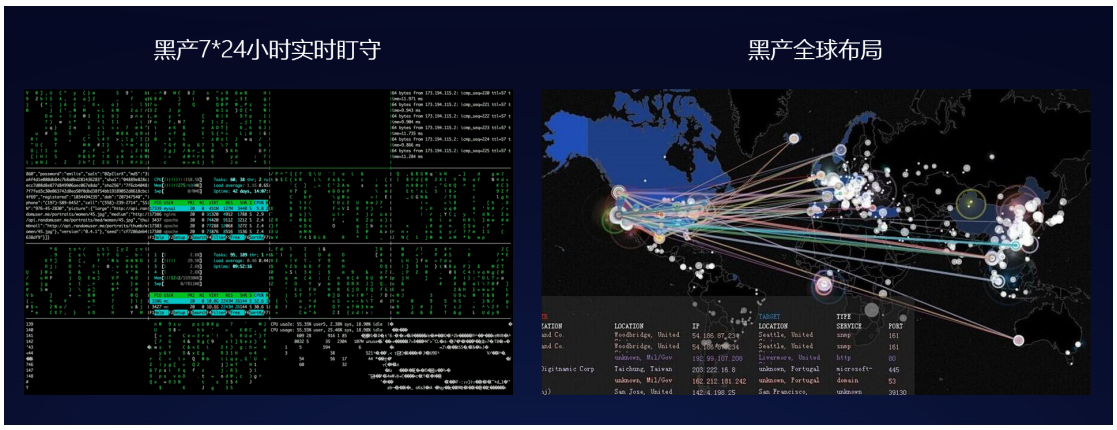

图 11 黑产态势

2.1.4 黑产链条完备，分工明确

黑产上下游分工明确，形成了产业链。

图 12 黑产链条示例

2.1.4.1 黑产情报

对于当今组织化规模化越来越强的黑灰产团伙来说，挖掘攻击情报往往是获利的第一步，团伙中会有专门角色负责欺诈线报收集，把相关活动的时间范围、收益变现形式等信息准确、及时地在团伙内传

达清楚,线报人员获取情报的来源通常包含黑灰产论坛、信息分享QQ/微信群、电报群等。

信息获取后，就会有专门的业务渗透人员和脚本人员，了解分析清楚产品逻辑、必需资源和必要工具/脚本，厘清活动性质，如是新账号首单还是老账号拉活，是否涉及地域性，是否涉及绑卡等。进而基于前期分析后做出相关操作决策，如充钱，屯号等，以确保在活动开始前，做好准备。

2.1.4.2 核心资源获取与基础工具

巧妇难为无米之炊，单个账号能薅的羊毛通常会有产品限制，且 IP 和设备相关限制也是企业做风控的基础依据，所以为了能有批量收益，提前准备好海量资源是重中之重。

（1）账号

欺诈账号的来源有两个，一个是注册，一个是盗号。

批量恶意注册需要批量手机号短信验证码，此类黑产分为几代：

第一代：虚拟运营商手机号，即 170、171 开头的手机号。虚商从 2013 年发展至今，已有阿里、京东、苏宁等几十家机构拿到了虚拟运营商牌照。虚拟卡主要应用在临时场景，办卡门槛较低，因此受到了黑灰产网络欺诈的青睐。尽管其成为较常见的黑产手机号来源，但因为识别简单，防御起来门槛较低。

第二代：海外、传统运营商流出的黑手机卡。其往往是处于欠费半停机状态（只能收短信）或 0 月租的号码。当前，越来越多的企业启用语音验证码，也出现不少质量较高的手机卡，可以接听语音。这类传统黑卡需要配合历史作恶黑库来识别，防御门槛稍高，但效率存疑。

第三代：注册时使用的手机资源还需要提升接码效率，猫池+卡池的工具组合应运而生，猫池负责解码接短信，模拟正常手机的功能，卡池为猫池提供足够卡源，实现全天无人工值守自动随机换卡，猫池和卡池的关系有点像步枪和弹夹。该类方案成熟度较高，但成本较高。

第四代：运用一、二代的卡资源，再结合三代的工具，就组成了一体化的接码平台。一体化接码平台可承接各类运营商的手机号，同时提供专门客户端和高并发 API 接口，采用会员充值制，给黑灰产注册欺诈提供了强有力的支持。

第二个号码来源是盗号，此类黑产同样也在不断变迁：

第一代：木马方式盗号。该产业链有明确的分工，有人设计木马程序，有人专门传播/控制木马，有人收集梳理盗取的号码库，有人负责有价值的号码变现，即“洗库”。随着移动互联网大潮来临，很多厂商转型移动端。

第二代：号码及票据方式盗号。当企业级服务的 web 端存在时，可利用 XSS 或 CSRF 等漏洞，使得 cookie 中登录票据等私密信息的泄漏。在此类登录票据的有效期内，黑灰产可以利用此号码+票据进

行盗号，并引发出多次漏洞、蠕虫恶意传播事件。

第三代：前述两代是通过漏洞攻防安全技术作恶，技术门槛高，且对抗激烈。第三代盗号，瞄准了安全链条中最薄弱的一环，人。当前，基于社会工程学的攻击层出不穷，花样迭出。盗号是社工的重灾区，无论是批量弱密码扫号还是仿造得惟妙惟肖的钓鱼网站盗号，都是成本较低且很难彻底防御的攻击手段。

第四代：撞库盗号。黑灰产利用漏洞进行拖库，结合人性弱点（每个人平均能记住的密码不超过 3 个，复用情况普遍），用手里的“库”进行暴力“撞库”成了盗号的最省力方式，且性价比很高。

（2）设备

用户设备是承载账号的硬件载体，模拟/更换设备是黑灰产的基本方法，分为以下几代：

第一代：基于协议破解的假设备。成本较低，但因无设备唯一标识，因此防御方法也相对简单。通过设计设备唯一标识并进行检测，即可事半功倍。防御重点主要在运营维护客户端版本，兼容历史包袱。

第二代：安卓手机模拟器。一台 PC 往往可以同时执行 15-20 个模拟器。但该类方法门槛较低，可通过设备风险识别等手段可以较轻松识别。

第三代：真手机设备篡改或多开。该类方法基于原生移动环境，通过改机工具来实现模拟换设备操作。当今风控厂商的设备风险识别

也基本可以识别到，门槛稍高。

第四代：设备农场。随着羊毛党收益等不断上涨，可以支持到黑灰产启用成本更高的作恶方式，真机设备农场就是典型案例，也即俗称的手机墙。黑灰产从二手货交易平台回收旧手机，统一刷机，结合三代的改机工具，批量作恶，给企业风控带来较大威胁。

（3） IP

用户 IP 是网络接入载体，更换 IP 也是黑灰产必修课，手法分为以下几代：

第一代：ADSL 拨号。通过一根网线反复拨号，可以拨到多个相邻 IP 地址段，但无法跨越更大的网段和地域。所以 ADSL 拨号虽然容易上手，但防御较容易。

第二代：VPN 代理。较传统的代理架设模式，VPN 私密性较好，部分场景用于翻墙，可模仿跨城市的 IP 网段，用初级黑库+行为逻辑可以进行防御。

第三代：专门用于黑灰产的代理服务器。包含扫描代理、肉鸡代理、私搭代理平台服务等形式，IP 和城市数量较多。但因为鱼龙混杂，良莠不齐，攻击维护成本较高。

第四代：VPS 混拨。远程在 VPS 云服务器上架设多根网线，软件实现多线混拨，远程控制 VPS 服务器拨号可以跨城市，速度快，稳定性好，稍有规模的 VPS 拨号厂商，可支持几十个省份上百个城市地域

几十万 IP 的混拨更换，是目前黑灰产使用的主流换 IP 模式。

2.1.4.3 黑产业务工具

- 打码平台

打码平台，验证码是人机识别的标配，那么自动破解验证码的打码平台服务也成了必需品，从人工打码到深度学习自动识别破解验证码，效率和收益都在不断提升。

- 脚本

按键精灵+脚本，看似简单，但实际上是流程控制的核心，根据产品业务流程，设计脚本，通过按键精灵来模拟用户操作，成本低又能加入各种随机因素保证模拟质量。

- 工具

资源与基础工具，组成各作恶场景的业务工具，如撞库工具、引流脚本、羊毛软件，把所有核心资源串联起来，提升作恶效率。

- 模拟器

模拟器是能在电脑上模拟手机操作系统，并能安装、使用、卸载手机应用的软件，它能让你在电脑上模拟手机整个操作的过程。

- 改机软件

改机软件是一种可以安装在移动端设备上的 app，能够修改包括手机型号、串码、IMEI、GPS 定位、MAC 地址、无线名称、手机号等

在内的设备信息，通过不断刷新伪造设备指纹，可以达成欺骗厂商设备检测的目的，使一部手机可以虚拟裂变为多部手机，极大地降低了黑灰产在移动端设备上的成本。

- 猫池

猫池是将相当数量的Modem使用特殊的拨号请求接入设备连接在一起，可以同时接受多个用户拨号连接的设备。恶意用户可以利用一台猫池使用多张手机号进行业务操作。

2.1.4.4 变现及套利

通过提现和各种实物或虚拟价值套利方式，进行变现，黑产团队里的“变现师傅”是不可或缺的重要角色，有多少资源人脉，能找到多高价格收的平台渠道，决定了黑产团队的经济命脉。

2.2 欺诈在移动业务中的趋势和特点

近年来，我国移动互联网用户增长迅速，智能终端的渗透率不断提升。据工业和信息化部显示，截止到 2018 年底，我国 4G 用户的数量已经超过 11 亿,移动终端已经成为最为重要的流量入口。同时，移动终端的功能不断增强，减少了用户操作的复杂度，使得用户随时随地获得需要的服务。然而，移动端服务在给用户带来更大便利性的同时，也给欺诈行为带来了更大的空间。目前，数字金融和电子商务中的欺诈行为中来源于移动端的占比逐年递增。 从今年来欺诈发生

的统计情况来看，数字金融和电子商务除了面临传统模式会遇到的业务风险、信用风险外，还增加了许多新的风险，如操作风险、信息风险，技术风险等，具体而言，我国移动数字金融和电子商务面临的欺诈风险包括以下几个新的特点：

2.2.1 行为模式：“被动”变为“主动”

移动互联网突破了传统服务行业的界限——将商家、用户、平台方、电信运营商，第三方支付企业等更多的角色投入到服务中来，从而实现提供更方便快捷、低成本、低门滥、不受时间地域限制的金融及购物服务。然而，也正是由于多方参与，当出现信息泄露时，不法分子更容易将这些相互关联的信息进行组合，从而拼凑出更为完整的“用户画像”。

另一方面，由于手机终端日益成为生活和工作的“必需品”，其时时刻刻与用户保持着联系。因此，不法分子也拥有了更容易、更多样的“主动”接近用户的渠道——而不再被动等待用户“上钩”。 尤其是当不法分子掌握了用户的多种信息，通过多种渠道同时针对某个个体实施欺诈行为时，更使得普通用户难以分别真假。从结果来看，近年来，被欺诈人群也逐渐呈现年轻化特征。传统上，易受骗人群多为中老年人，而随着移动互联网业务的不断普及，欺诈分子更倾向“主动”触及年轻人，同时通过同渠道进行欺诈，使得诈骗流程变得更短、更快。

可见，移动互联网服务的发展一方面为用户带来了便捷，另一方

面也让欺诈分子不再“守株待兔”，而是采用定制化的欺诈方案“主动出击”，提高了不法分子实施欺诈活动的效率，给我国数字经济的发展带来更大的负面影响。

2.2.2 安全漏洞：“碎片”变为“系统”

安全漏洞是在应用中软件、协议的具体实现或系统安全策略上存在的缺陷，从而可以使攻击者能够在未授权的情况下访问或破坏系统。移动互联网时代，信息安全的风险很大程度来源于系统技术演化中出现的不适用性。移动互联网的信息安全风险主要出现在移动应用程序（App）上。对于传统的信息化系统的信息安全问题，近年来已经得到了有效的管理。但是，随着移动应用和云计算的出现、新的信息安全问题很难有效检测并予以控制。包括身份识别与合理授权存在难度、进程管理等问题使信息安全面临更加严峻的形势。根据 **Web 安全和渗透测试厂商 Cenzic 公司最新发表的报告**统计，有将近 96%的移动应用上存在着安全漏洞问题。

上述安全风险逐渐从“碎片化”的单点风险转变为影响全局的“系统性”风险。其一旦被欺诈黑/灰产行业利用，将对个人信息、企业安全等造成重大影响。

2.2.3 商业逻辑：“孤岛”变为“融合”

随着我国“互联网+”推动不断深入，在大量社会资本的追捧下，新的模式创新层出不穷。这些创新大大推动了我国移动互联网行业的

发展，使得新的移动互联网业务迅速下沉普及。然而，与此同时，整个行业也存在着“为了融资而创新”、“为了讲故事而创新”的苗头，对于业务创新过于热衷，而没有对新模式可能带来的业务风险给与足够的重视。需要看到，移动互联网服务不管在模式上进行怎样的创新，其本质还是服务，所以它也会面临传统服务模式同样具有的风险。其中比较突出是信用风险、业务风险和误操作风险。

信用风险主要由于在我国的移动互联网服务中，为了吸引客户，减少交易步骤可能带来的用户流失，很多企业都有意或无意地放宽了对于用户信用的评估，从而带来的业务履行或服务中可能存在的欺诈风险。移动互联网服务中信用欠缺问题主要体现在两个方面，一个是自身存在的信用体系构建问题，另一个是对用户以及业务的信用程度缺乏把握。我国移动互联网服务的信用风险相对较高，这是因为在移动网络中，数据遭到篡改的风险很大，真实性和可靠性降低，如网上出现的各种“代刷信用”、“伪装用户”、“修改评价”等现象，甚至已经发展成为地下产业链，这给移动数字金融和电子商务活动的开展带来了十分重大的风险。

业务风险主要是由于移动互联网服务的商务模式往往采用“羊毛出在猪身上”的模式，即以获取客户为目的，而不是以从服务中直接获取收益为主要目的。这就导致很多移动互联网企业以“撒钱”的方式，通过赠品、大额折扣、甚至现金等方式争夺用户。这种方式一方面存在严重的业务风险，比如“羊毛党”、“黄牛党”留下了大量漏洞，造成宝贵的社会风投资本被大量浪费；另一方面，当企业长期无法“自

我造血”而陷入困境时甚至“跑路”时，也给正常购买服务的用户带来损失。这种情形在 P2P、共享服务等领域频频出现，屡见不鲜，给我国移动互联网业务的发展和口碑带来不影响。

误操作风险主要是用于很多用户对于新的服务不了解而出现的失误，以及被有意误导造成操作失误，从而引起的损失。在移动互联网服务中，用户往往会实名认证并进行账户绑定，这样可以让支付过程获得很大的便利。但与此同时，过于简洁的流程也使得用户更容易被误导进行并非本意的交易。另一方面，由于我国移动互联网金融尚未形成一个统一、规范的操作流程，不同的金融软件如手机银行、第三方支付等其操作流程存在一定的差异性。很多不法分子就是利用这一现状，通过隐藏在应用软件中的插件误导用户误操作，从而在实现非法牟利的目的。

移动互联网的模式创新本质是提高效率，降低成本，规模化运营以获得收益。另一面，互联网+“业务”的深度融合也进一步加剧了上述风险的融合，给黑/灰产在欺诈活动更大的空间和便利。

2.2.4 变现逻辑：“量变”变为“质变”

PC 互联网时代，早期的互联网场景更多的是线上的展示（广告类）。所以黑灰产的欺诈行为是通过控制的个人电脑做为肉鸡来进行 Ddos、刷广告、安装流氓软件等变现。在这个时代一台台实际的物理电脑就是黑灰产的核心资源，资源意味着变现能力。而在移动互联网时代，互联网场景变成了场景和个体的连接。所以黑灰产的欺诈行为

也变成了通过大量的手机号在各个场景下注册虚假账号，并通过这些账号在业务场景中变现。利用这些虚假账号进行恶意注册，黑灰产可以在多个平台实现变现，如：电商平台薅羊毛，金融平台骗贷，电信诈骗，短视频平台刷量等变现方式。黑产可控的设备和资源的海量增长，使得黑产的变现能力从“量变”到“质变”。

2.2.5 迭代速度：“缓慢”变为“迅速”

相较于正常软件，黑灰产工具软件具有更强的版本更新迭代速度。除了增加新功能，修复 Bug 以外，频繁的版本更新更是黑灰产从业人员跟业务安全团队攻防对抗加剧的体现。一个比较典型的场景：一款针对 X 产商的工具软件发布一段时间之后，通过业务侧的数据和模型，X 产商的业务安全团队感知到了由于工具软件产生的异常，并通过修复漏洞，改进监测模型等方式使工具失效；而工具软件的作者则需要重新找到新的突破口，然后发布新版本。

三、移动数字金融和电子商务领域的反欺诈方案

3.1 现有反欺诈方案面临的挑战

当前，有很多厂商提供反欺诈解决方案。同时，大多互联网、“互联网+”企业都部署有不同类型的欺诈检测系统，但很多方案尚不够完善，主要体现在以下几个方面：

（1）防御能力单薄

- 通过单个技术方式进行欺诈识别，比如单纯依赖黑名单或简单的人工规则、单点布控等，缺少全流程的反欺诈方法（如设备指纹、验证码、注册保护、登录保护），无法形成从设备启动到用户行为各个环节的纵深防御。
- 反欺诈功能薄弱，只能防御部分欺诈形式，而不能全面防御羊毛党、虚假用户裂变、渠道流量作弊、广告导流、刷单、内容爬取等欺诈形式。

（2）防御时效性差

- 比如依赖 T+1 离线挖掘，欺诈发生后才能发现，无法实时、在线阻止欺诈损失的发生，同时策略生效周期长。

（3） 防御进化慢

- 缺乏策略迭代闭环，没有形成攻防研究 – 策略设计 – 策略研发 – 验证 – 运营 – 案例分析 – 策略更新的进化闭环。
- 无自学习机制，难以发现潜在的和新型的欺诈模式。

（4） 过分依赖设备显性 ID

- 过分依赖“显性 ID”，如 IMEI、MAC、IMSI、SN 等，以此为识别设备唯一性的标准。
- 这些 ID 都可以通过改码软件或虚拟机等手段轻易的进行变换/模拟，从而利用有限的造假设备就可以伪装出大量的移动设备。

（5） 黑名单库准确性低

- 数据服务商受利益驱动，希望扩大数据体量，赚取更多查询费用。
- 整个行业对于黑名单没有一个明确的规范标准。A 的黑名单库，未必适合 B 公司。
- 黑名单数据僵化，因其不可证伪性，更新迭代速度慢，其反欺诈效果越来越差。利用黑名单反欺诈，不能满足金融行业既要做防范风险，又要避免误伤无辜的需求。

3.2 全栈式实时反欺诈方案

完善的反欺诈方案应该既是全面的，也是实时的。全面的反欺诈方案应该包括：全场景识别体系、全路径实时布控体系、全方位策略体系和全流程运营体系。这样的方案可以全面覆盖全场景多维度，维护企业健康生态，为企业持续运营增长保驾护航，免受黑产欺诈干扰。实现营销活动反作弊、内容防盗爬、渠道流量反作弊、刷榜刷单防护、虚假用户裂变识别、支付风控等全方位安全保护。

图 13 全栈实时反欺诈方案

3.2.1 全场景识别体系

智能识别不同行业复杂多样欺诈场景，包括移动数字金融行业的伪冒转账、信用卡银行卡盗刷、反洗钱、营销活动欺诈，电子商务的刷榜刷单、商品评论广告导流、机器秒杀特价商品、虚假交易套利，虚假用户裂变等众多场景。

3.2.2 全路径实时布控体系

全路径纵深布局，从设备启动到用户注册、登陆，再到浏览、签到、领券、分享、送礼、邀请好友、评论、下单、提现等业务行为每一个环节实时识别风险，在每一个风险点都精确部署专家防御策略集，做到全局实时风控。

图 14 全路径实时布控体系

（1）启动

启动阶段根据设备风险 SDK 采集数据判断设备是否存在风险，具体包括篡改设备、虚拟机、设备农场、积分墙、设备风险分、云手机等。

（2）注册

注册保护通过作弊设备识别，风险设备聚类，风险设备库，风险 IP 库，风险手机号等识别批量接码注册、批量脚本注册、注册机注册、模拟器注册、代理 IP 注册、异常手机号注册等异常注册行为。

（3）登录

登录保护通过作弊设备识别，风险设备聚类，异地登录检测，风险设备库，风险 IP 库等识别撞库尝试、撞库盗号、洗号登录、晒号登录、养号登录、模拟器登录，篡改设备登录，异常手机号登录、代理 IP 登录等异常登录行为。

（4）业务行为

业务行为保护主要判定机器刷积分、代下载、自动化抢红包、机器阅读、自动刷单、自动点赞、盗刷卡、机器秒杀等异常行为。

3.2.3 全方位策略体系

构建全方位多维度立体策略集合，包括关联图谱、行为时序、属性聚熵、资源离散、目标聚集、时域网络，检测设备、行为、关联风险三个层面，实时识别虚拟机、篡改设备、设备农场、积分墙设备、云手机等风险设备；实时拦截机器注册、撞库攻击、机器秒杀、机器薅羊毛、虚假拉新、虚假裂变、恶意抢红包、恶意刷榜刷单等风险操作；同时结合时域网络，采用无监督机器学习模型，递归调度 PageRank 等风险传播算法进行黑产社群发现，多方位有效精准识别欺诈团伙和高危群体。

3.2.4 全流程运营体系

设立行业级黑产研究院，持续跟踪黑产动态，同时部署顶级策略团队，打造全流程持续循环运营体系，从“案例分析”、“攻防研究”、“策略设计”、“策略研发”、“策略验证”、“策略上线”到“策略运营”，实现攻防策略的“闭环迭代，持续进化”。

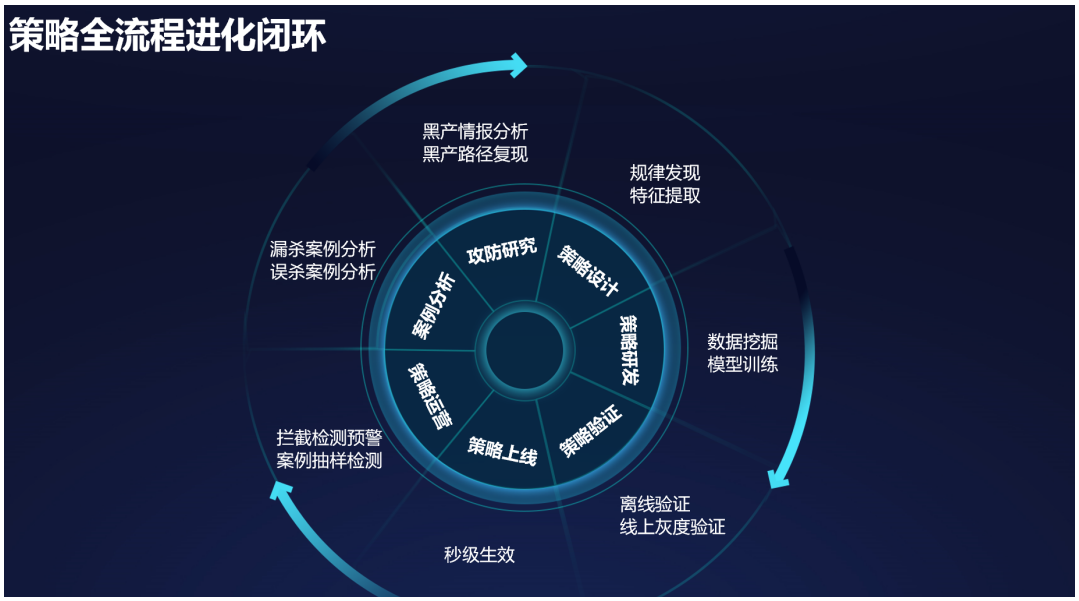

图 15 全流程闭环策略体系

3.3 移动设备唯一性甄别实时反欺诈方案

移动设备唯一性甄别实时反欺诈方案，是指利用移动设备标识如“可信 ID”，为移动开发者实时提供设备真实性&唯一性的甄别服务。通过有效的反作弊措施，鉴别虚假数据，提升运营数据质量，从而有效杜绝灰色产业链的侵蚀。

其采用具备编译器级别的反编译方法，使用数据动态加密，进行代码混淆加密，以及运行环境识别，防止 SDK 被破解以及传输数据被伪造。

架构上采用多地采集中心和双存储中心，能够容灾备份，保证数据安全。

数据采用流式处理和分布式处理，毫秒内返回设备及应用状态结果，保证客户数据实时性。

平台架构分为系统应用层、消息枢纽层、数据计算层、数据存储层，以及监控管理系统。系统应用层包括数据采集上报、客户服务系统;消息枢纽层使用 Zookeeper+Kafaka;数据计算层包括离线的大数据计算和实时的业务数据计算；数据存储层包括原始日志、中间及结果型数据。

主要反欺诈应用场景有：O2O 防薅羊毛、反作弊验证、虚假流量识别、裂变红包防薅、虚假账号识别、数据防爬虫、伪基站校验识别、识别恶意卸载等。

3.3.1 账号识别及保护反欺诈方案

（1）虚假账号识别/识别欺诈小号：依靠可信 ID 与客户自定义 ID 关联服务，实时鉴别欺诈小号，方便开发者制定设备与账号的反欺诈规则。精准识别虚拟机、模拟器、改码设备生成的账号，对虚假账号进行有效清洗，提升运维质量。

（2）账号保护：防范撞库攻击、暴力破解、账号盗号等恶意行为带来的商业损失。

3.3.2 营销活动反欺诈方案

（1）防羊毛党：实时判断设备真伪，过滤机器注册及重复领取行为，封堵推广资金被“薅羊毛”，避免推广资金被盗取。

（2）裂变红包防薅：识别裂变红包营销中的作弊行为，通过设备可信 ID 锁定背后的用户，毫秒级鉴别机刷红包及重复领取行为，

只给有效用户发放红包福利。

（3）虚假流量识别：提供准确的渠道效果监测服务，毫秒级实时数据反馈、可有效识别虚拟机、重复、封信、召回等多种行为，有效鉴别 App 营销推广中的虚假流量。

（4）渠道推广反作弊：提供有效的渠道推广反作弊服务，毫秒级实时数据反馈、可有效识别虚拟机、重复、封信、召回等多种行为，有效鉴别 App 营销推广中的虚假流量。

（5）反作弊验证：实时验证移动运营推广中的数据质量，基于设备唯一性识别的技术，有效鉴别改码手机、模拟器、虚拟机等生成的虚假用户及流量，提升防作弊的风控等级。

3.3.3 网络安全/提供风控方案

（1）数据防爬虫：高效识别并及时遏止搜索场景中被爬虫盗取内容或数据的行为，保护商业敏感数据，减少企业运营风险。

（2）识别恶意卸载：提供移动设备 root 查询服务，识别恶意卸载行为，降低运维风险。

（3）校验识别伪基站：保护短信通道安全，防范短信接口被恶意调用、滥用带来的利益损失。

3.3.4 互联网金融反欺诈方案

（1）欺诈黑名单：基于用户历史互金行为和风险规则分析来识别和定义黑名单。历史互金行为数据由互联网金融企业采集，并由第

三方服务机构进行统一维护。第三方服务机构提供对外查询接口，从而帮助其他互金企业方法来自黑名单用户的风险。另外，拥有自主数据源的第三方服务机构，也可以基于自身数据构建风险规则和判定模型，建立欺诈黑名单。

（2）地理位置验真：移动大数据服务机构可以根据用户上报的GPS、WiFi、基站信息，筛选最近三个月内工作时间和休息时间最常出现地点，判断用户所在地址。用户地址与互金用户登记地址不符的，可能存在欺诈风险。另外，地理位置POI信息也可以判断出用户差旅频次、购物偏好和酒店住宿偏好等。这些信息有助于帮助互金企业进行风险识别和金融授信。

（3）风险因子：移动大数据服务机构通过对最底层的数据建模、训练、测试、优化后提炼出的一些与好逾期行为潜在相关的变量，经过脱敏加工处理之后输出计算结果。互金企业可以通过实时查询的方式判定用户的风险水平。这些因子包括但不限于：

- **设备网络关系指数**
  - 手机应用稳定使用特征
  - 换机换号行为特征
- **金融借贷行为因子**
  - 小额贷款，个人小额贷款类应用的使用时长、使用时间、频度等，综合计算得出的强度指数。
  - 现金借贷、现金贷款类应用的使用时长、使用时间、频度等，综合计算得出的强度指数。

- 信用卡套现、信用卡贷款类应用的使用时长、使用时间、频度等，综合计算得出的强度指数。

- **生活服务行为因子**
  - 使用餐厅推荐、餐厅点评类应用的行为特征强度指数，基于当前在装使用的相关应用的数据，综合计算得出。
  - 使用餐厅外卖、外卖订餐类应用的行为特征强度指数，基于当前在装使用的相关应用的数据，综合计算得出。
  - 使用餐厅优惠券、餐厅代金券类应用的行为特征强度指数，基于当前在装使用的相关应用的数据，综合计算得出。
- **金融支付行为因子**
  - 在装使用的手机支付类应用中信用支付，额度支付类应用的使用时长、使用时间、频度等，综合计算得出的强度指数。
  - 在装使用的手机支付类应用中商家收款，支付软件商家版类应用的使用时长、使用时间、频度等，综合计算得出的强度指数。

四、反欺诈的技术与效果评估

4.1 反欺诈技术体系架构

整体流程如下图：

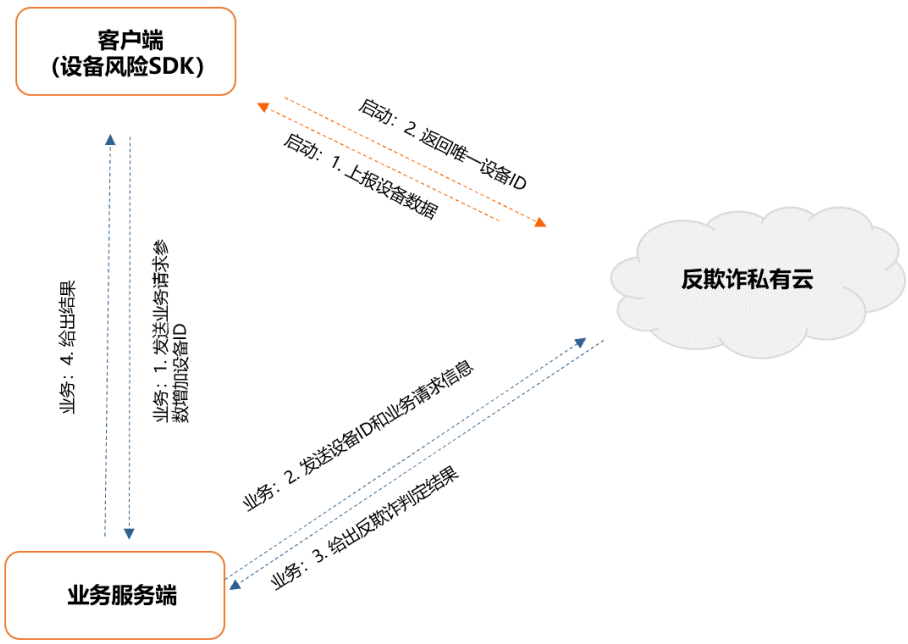

图 16 反欺诈技术流程体系

客户端需要嵌入设备风险 SDK，在客户启动阶段调用反欺诈私有云，在注册、登录等业务行为事件再调用反欺诈云，详细流程如下：

（1）设备启动阶段

- 客户端上报设备数据到反欺诈云
- 反欺诈云返回唯一设备 ID，并计算设备画像特征

（2）业务事件阶段

- 发送业务请求参数（包括设备 ID 等其他业务参数）到

业务服务端

- 业务服务端发送请求参数（包括设备 ID 等其他参数）到反欺诈云
- 反欺诈云返回判断结果给业务服务端
- 业务服务端根据结果给出处置建议，并返回客户端

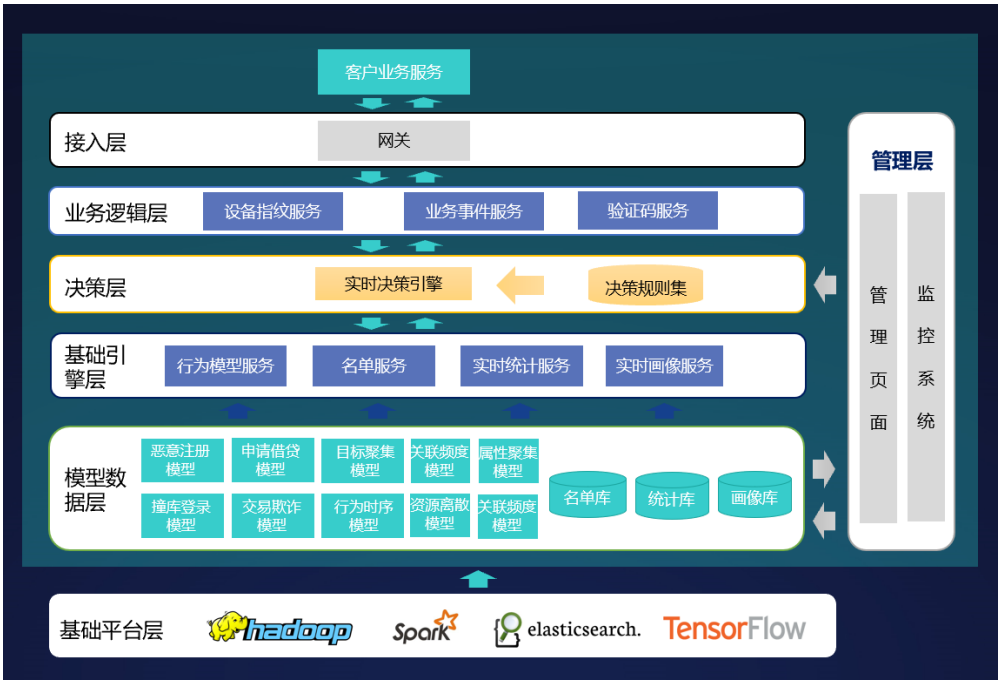

图 17 反欺诈云架构

每个层的功能如下：

4.1.1 接入层

主要进行接口安全检查、流量控制、参数校验、过载保护等功能。一般提供 HTTP 接口，POST 请求方式，返回结果的格式一般为 JSON 格式。

4.1.2 业务逻辑层

主要包括业务逻辑的处理，具体模块包括设备指纹服务、业务事件服务、验证码服务。

4.1.3 决策层

决策层包括决策引擎（AE），加载策略集合，根据输入和基础引擎的结果进行判决，给出最后的判定结果。

4.1.4 基础引擎层

基础引擎层包含行为模型服务、名单服务、实时统计服务、实时画像服务，计算各种基础特征。

（1）行为模型服务

行为模型服务是一组基础引擎。它提供了包括监督学习、非监督学习、时域关联网络等多类机器学习在内的模型加载与运行服务。

监督学习模型包括：

- 经典的 LR、SVM、GBDT、RandomForest 等
- 深度学习模型，RNN，GAN

非监督学习模型主要是一些异常检测模型：

- GMM 异常检测、离散度模型等

时域关联网络：

- 通过在时间和地域上通过各个实体之间关系进行团伙挖掘

（2）名单服务

名单服务提供配置黑白名单的功能，包括设备、IP、账号等黑白名单。

（3）实时统计服务

配置相关统计指标，包括统计窗口、统计维度、统计周期等，实时统计服务进行实时统计所有配置的指标。

（4）实时画像服务

根据配置相关画像变量，进行实时计算。

4.1.5 模型数据层

提供相关模型和数据，模型包括恶意注册模型、撞库登录模型、申请借贷模型、交易欺诈模型、目标聚集模型、行为时序模型、关联频度模型、资源离散模型、属性聚熵模型等；数据包括名单库、统计库、画像库。

4.1.6 基础平台层

数据平台层主要包括大数据分析和建模平台，具体包括 Hadoop、Spark、Elasticsearch、Tensorflow。用于模型训练和数据分析。

4.1.7 管理层

包括具体规则数据管理、系统监控等。

4.2 反欺诈技术详解

4.2.1 反欺诈情报体系

构建反欺诈情报体系，从数据收集、情报分析、风险量化几个维度出发，以攻击者的视角了解自身业务风险状况。

a.黑灰产情报数据收集：通过对黑灰产的沟通渠道、交易平台、核心资源、攻击流量、攻击工具等多个维度进行监控，收集黑灰产情报数据；

b.情报分析、运营：对收集到的黑灰产数据进行清洗、分析、统计和挖掘转换成有价值的情报信息，了解黑灰产最新动向；对黑灰产工具进行逆向分享，发现其利用的业务逻辑漏洞。

4.2.2 设备指纹技术

设备指纹主要在客户端主动地收集与设备相关的信息和特征，通过对这些特征的识别判断识别的风险程度，同时生成唯一 ID，用于标识设备。

设备指纹主要有五个方面的作用：

图 18 设备指纹的作用

（1）设备的有效性

甄别设备的有效性，判断是否为虚拟设备。虚拟机（模拟器）原意是用来给 Android 开发人员调试 App 用的，开发人员可以通过不同的设置的虚拟机来适配不同的屏幕尺寸、分辨率，也可以用来模拟电话、电话、定位等。图所示即 Android 原生的虚拟机。除了原生的虚拟机，市面上还有很多专门开发虚拟机的厂商，如夜神、逍遥等等，这类虚拟机本意是给玩家玩 Android 平台的游戏用的，比如说有的游戏可能暂时只在 Android 平台上线了而用户又没有 Android 的设备，或者因为使用虚拟机可以通过鼠标键盘操作，会比用触摸屏操作方便，所以有部分玩家会选择虚拟机来玩游戏。

图 19 虚拟机示例

由于虚拟机可以很容易地抹掉设备的数据，修改设备的信息，因此部分黑产会利用虚拟机来作恶。某些场景下，黑产需要获得大量的设备，例如部分 App 对新用户发放优惠券，但一个设备只允许领取一次，如果黑产想要获取大量的优惠券，就需要搞定大量的设备，而虚拟机能以较低的成本为黑产提供无限量的设备。

一般来说黑产们会直接使用市面上的虚拟机，他们一般没有能力也不会自己去实现一个虚拟机。在判断设备是否为虚拟机时，应重点考虑两方面。一方面针对市面上已有的虚拟机特征进行检测，例如部分虚拟机会独有特定文件，另一方面，应基于虚拟机的原理挖掘的特征进行检测。

（2）设备属性检测

检测设备属性是否异常，原始属性是否被篡改。

篡改指修改设备信息，在 Android 平台上，篡改工具一般可修改以下信息：手机品牌、机型、手机号、IMEI、IMSI、Android ID、MAC

地址、GPS 位置信息。部分篡改工具甚至可以修改 cpu 信息、编译信息等。从原理上讲，Android 提供了 API 的信息都可以被篡改。与 Android 类似， iOS 平台上，一般篡改 IDFA、IDFV、UUID、MAC 地址等信息。

篡改一般情况下需要 root 手机，此外有可能还需要安装特定的框架，如 Xposed。iOS 的篡改需要越狱。

图 20 安卓和苹果设备信息篡改示例

篡改是获取大量虚拟设备的有效方法。通过篡改一个手机的设备信息，就能使其变成海量虚拟设备。当前市面上的很多篡改工具都提供“一键新机”的功能，只需要按一个按钮就能一键修改掉设备的关键信息，大大降低了黑产作恶的成本。

目前市面上 Android 篡改工具种类较多，如 008、Czero、海鱼魔

器等，iOS 如 NZT。。虽然篡改工具很多，但其篡改的原理都是类似的，我们一方面会检测是否有安装这类篡改 App，另一方面会识别一些关键函数是否被篡改。

（3）作弊环境检测

主要检测是否有作弊环境，包括恶意作弊 APP 等。黑产常用的作弊包括多开软件、VPN 等。

- 多开软件

多开原意是帮助用户实现一机多号，例如有多个微信号的用户，可以利用多开 App 同时登录多个微信号。目前市面上的多开软件多达数十上百款，部分手机 rom 中还自带多开的功能。

图 21 多开软件示例

利用多开软件，一方面可以达到在一个手机上注册、登录多个账号，以达到薅羊毛的目的；另一方面利用多开可以同时登录多个账号

来发送引流文字、图片，减少了切换账号的麻烦。

Android 市面上的多开有两种形式。第一种的代表是“多开分身”，其会制作并让用户安装新 App，其作用是去加载被分身的那个 App。第二种是“平行空间”或“双开助手”，该类 App 多开不需要安装一个新 App，直接在多开 App 中加载被多开的 App 即可。但不管形式是怎样，其原理类似。

- VPN

VPN 可以简单理解为代理，网络请求通过 VPN 中转。利用 VPN 可以达到切换 IP 的目的，假如黑产一直在同一个 IP 下作恶，很容易被检测到，利用 VPN 切换 IP 可以避免掉这种 IP 的聚集性。目前市面上可以买到这类工具，也有一些免费的 VPN。

（4）环境变化检测

检测用户使用的设备环境是否发生变化。如，是否同一账号、是否同一设备，是否一个设备上有多个账号。当环境发生了变化，有可能出现了账号盗号或者是黑产在篡改设备的问题，这个时候，一般要加强进一步的验证，例如弹出验证码验证、活体识别验证、人工电话确认验证等，确保用户与系统安全。

（5）扩展服务

设备指纹采集的相关数据，可以与其他业务相结合，比如进行个性化推荐、可性度评分、设备聚类分析等。

4.2.3 实时决策引擎（规则引擎）技术

实时决策引擎，加载规则策略集合，根据输入和基础引擎的结果进行判决，给出最后的判定结果。主要是通过专家规则的方法进行识别，需要大量的专家知识经验来设计规则。

（1）功能设计

指管理员平台应该能够配置的功能(或者至少能够通过命令行工具配置，无需升级实时决策引擎)，它需要表达目前已有规则使用情况。

（2）策略动态配置

实时决策引擎可以针对不同的组织，不同的场景制定不同的策略。举例来讲，假设组织 1 有评论文本过滤、私信文本过滤和注册保护 3 类需求，组织 2 有评论文本过滤、注册保护 2 类需求。它们可能会形成如下的策略表格（表格 1），其中的 5 个策略可能各不相同。其中，(组织 1，评论)就构成了一个策略场景，其它场景类似。

表 5 策略动态配置示例

	组织 1	组织 2
评论过滤	策略 A	策略 B
私信过滤	策略 C	
注册保护	策略 E	策略 F

（3）RETE 算法

在规则执行中，实时决策引擎采用了高效的 RETE 并行推理算法。如下图所示，该算法保证的专家配置或自动生成规则的快速执行，数千条规则可以在 10 毫秒级别完成执行。

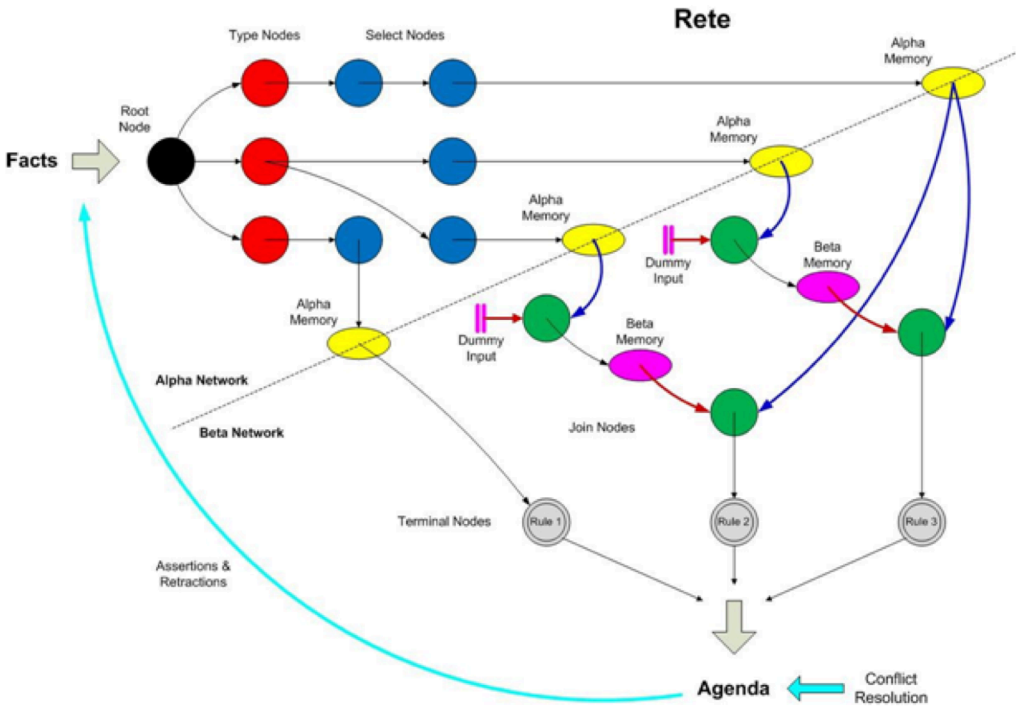

图 22 RETE 算法

RETE 算法可以对匹配阶段进行高效实现，主要包括鉴别网络和模式匹配过程两个方面。具体介绍可参考附录 B。

、规则引擎存在最大的问题在于只能依靠人为经验。这是一个十分耗费人力物力的过程，甚至还会出现经验错误而导致误判的状况。

4.2.4 知识图谱

引入具有应用价值的互联网系统的数据，在数据的基础上，进行

机器学习建模，形成账号、设备、IP 和历史行为的黑产知识图谱。当用户访问电渠系统时，基于黑产知识图谱，对用户身份进行风险防控。

黑产知识图谱如下所示：

图 23 知识图谱示例

其中包含手机号、IP、物理位置等多维度数据。

■ 手机号码维度

电渠的用户行为信息主要来源于订单、访问、支付等业务数据。但在很多业务场景中，如：新用户注册、营销推广等业务过程中，很多用户在电渠中没有任何业务数据或信息严重不足，很难基于行为进行风险分析。而用户除了访问电渠系统以外，在互联网的其它系统也有访问行为。为了消除“信息孤岛“，需要引入其它系统的风险数据，当用户访问电渠系统时，参考用户在其它系统的风险数据，对其进行综合风险评估。

■ IP 维度

用户访问电子渠道系统，所有的网络请求都会带有 IP 信息，因

此天然的成为访问者的身份标识。虽然 IP 地址极容易通过技术手段进行篡改，ip 数据的真实性难以界定，但是由于移动用户的特殊性，识别用户 IP 成为用户身份反欺诈的主要依据。

■ 地理信息维度

地理位置信息是指通过 GPS 定位或者基站定位的定位技术来获取手机或终端用户的位置信息（经纬度坐标），在电子地图上标出被定位对象的位置。通过结合 IP、设备等信息，可以有效识别用户的风险。

黑产知识图谱的测试流程如下：

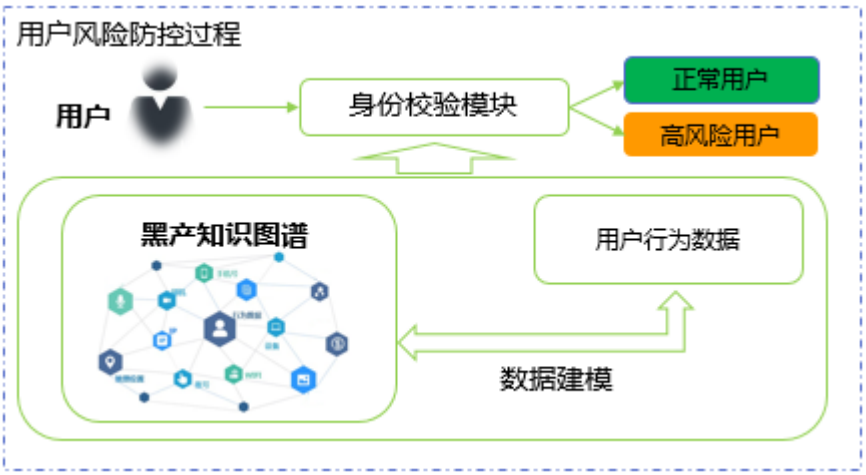

图 24 黑产知识图谱建模

4.2.5 有监督机器学习技术

有监督学习是当下应用最为广泛的反欺诈方法。有监督式学习是使用基础事实完成的,需要大量的有标签数据来训练模型，或者换句话说，我们事先知道样本的输出值应该是多少。监督学习的目标是学习一个函数，该函数在给定样本数据和期望输出的情况下，最接近于

数据中可观察到的输入和输出之间的关系。拿一个论坛的黑产广告帖子举例，假设你把 10000 个帖子已经由人工确认过黑产广告文章输入到了模型，模型通过对标题的识别，对文章内容句子的分割，关键词的识别等各种分析方法，找到了其中的内在关系，但却难以说明。

举例：

- 标题里有“美女视频”四个字的，有 70%的可能性是黑产广告；
- 出现了“激情表演”四个字的，有 80%的可能性是黑产广告；
- 内容里有“请加微信：xxxxxx”的，有 60%的可能性是黑产广告；
- 发布到 200 以上的论坛，有 85%的可能性是黑产广告；
- 回复率低于 10%的帖子，有 70%的可能性是黑产广告；

这里，百分比被称为权重。

当模型处理一封新的邮件，模型通过检测以上各个子项，并对每一个子项乘以权重后相加，就得出一个分数，得出：这封有 80%的可能性是垃圾邮件。

以上就是一个有监督学习抽象理解的过程，其中一个重要的步骤就是通过不断的迭代计算每一个子项应该被赋予的权重值。当权重值计算好后，就可以说这个模型训练好了。

有监督学习的好处也十分明显，它可以帮我们分析隐层关系。无需知道有监督学习的隐藏关系，每一个子项被赋予了多少权重，直接

就知道符合某个规则的是坏人。此外，有监督还有助于处理多维数据。由于规则是人凭经验产生了，当面对巨量的数据字段时，人显然是无法通过经验来直接操作的。此时，有监督就可以发挥用场了。

有监督也有一个明显的弊端，每一个模型都需要大量的训练数据，训练一个模型也需要较长的时间。常常出现你的模型还没有训练好，欺诈分子们就可能已经完成欺诈活动寻找下一个目标了。

4.2.6 无监督机器学习技术

无监督学习中最常见的任务是聚类，表示学习和密度估计。在所有这些情况下，我们希望了解我们数据的内在结构，而不使用显式提供的标签。一些常用算法包括 k 均值聚类、主成分分析和自动编码器。由于没有提供标签，因此没有具体的方法来比较大多数无监督学习方法中的模型性能。

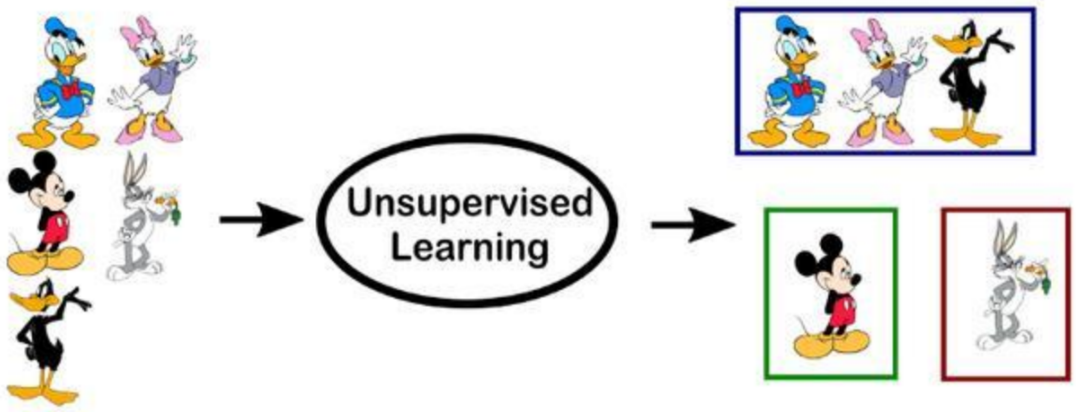

图 25 无监督学习

在反欺诈领域，无监督无需任何训练数据和标签，通过发现用户

的共性行为，以及用户和用户的关系来检测欺诈。

比如有这样一群用户注册事件，通过聚类发现其几个小群符合某些共性：例如：注册时间集中，都使用了某操作系统，某一个浏览器版本等。这个用户任何一个单独拿出来分析，看上去都是极其正常的用户，而如果其符合某种超乎寻常的一致性就十分可疑了。例如，一群人在凌晨 3 点-4 点，采用手机的都被 root，操作系统版本一致，注册了某产品，其 IP 的前 19 位相同，GPS 定位小于 1 公里，且注册后都使用美女头像，修改了性别均为女。如果一个人这么做，问题不大。而如果一群人这么做显然就是不正常的。

不过无监督也存在一个很大的弊端，就是时效性不够，往往是事后追踪，欺诈份子已经获利了结后才发现问题。

4.2.7 实时画像引擎技术

实时画像引擎技术实现了多环节联动联防，主要是为了实时发现欺诈问题，实时标记欺诈份子。

实时画像对每个事件都产生和计算画像特征并进行实时沉淀，同时给决策引擎提供画像数据，如下图：

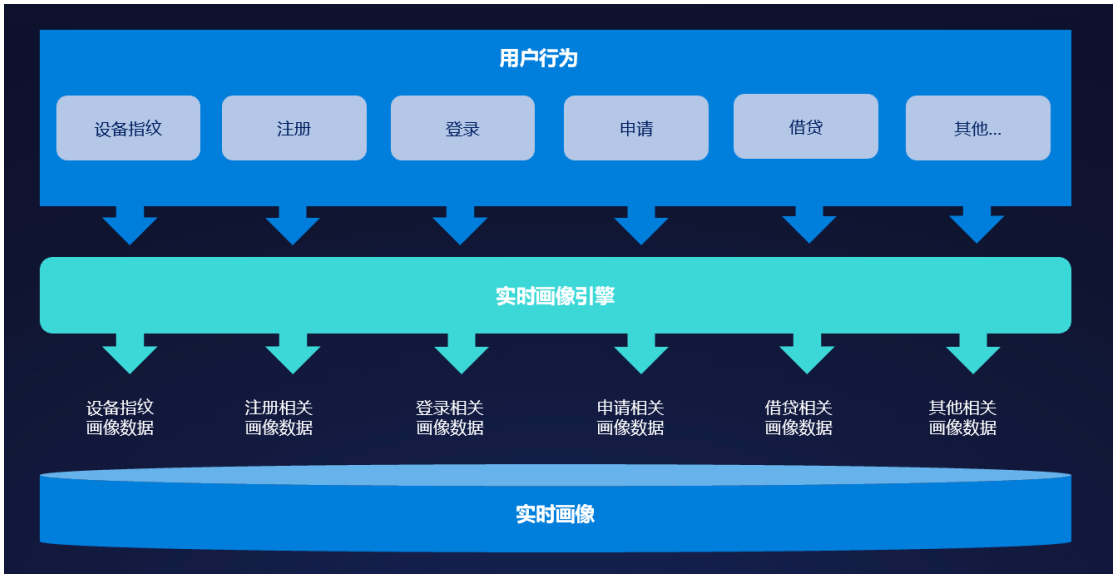

图 26 实时画像数据流转示意图

对于一个欺诈份子，如果有很强的特征能够迅速识别，直接可以使用简单的规则就可以识别，例如，使用设备农场进行注册，但是聪明的欺诈份子会伪装自己，让我们难以发现强特征，这个时候就需要实时画像技术，实时使用弱特征进行综合计算，比如以下一个例子：

一个用户在设备启动阶段，设备被 root 过了，这是个弱特征，我们不能简单判断是黑产，因为有些手机发烧友也经常 root 手机；

这个时候，这个用户进行了注册，头像使用了一个美女头像，我们不能认为是美女头像的人就是黑产，但是这也是一个弱特征，和上面特征结合起来，可疑度变高了，因为从概率上讲，手机发烧友一般技术宅男偏多，美女较少；

过了一段时间，这个用户又进行了异地登录，2 个小时类在三个城市都登录过了，如果单从这个行为也不能判断为黑产，因为 2 小时三个城市登录很有可能是在高铁上出差，但是结合前面两个事件，这个用户不正常的概率又变高了；

紧接着，这个用户在我们 app 评论区上发表了一个评论，“傈裱演，请 v❤️：xy！12-③”，这句话，使用 NLP 识别的结果，不是广告，而是疑似广告，疑似广告一般不会直接拦截，但是结合上面一系列的行为，可以最后判定，这个用户实际上是个黑产用户，正在使用机器进行黑产广告导流。

实时画像技术，可以很好解决类似这种问题，用户每一次操作中，都实时沉淀了行为特征，存入到了画像库中，并在每一次行为判断中，都使用了这些画像特征进行综合判断，从而实现实时的跨环节联动联防。

实时画像最主要的作用就是实时的沉淀与计算画像。

其具体实现的架构如下图所示。

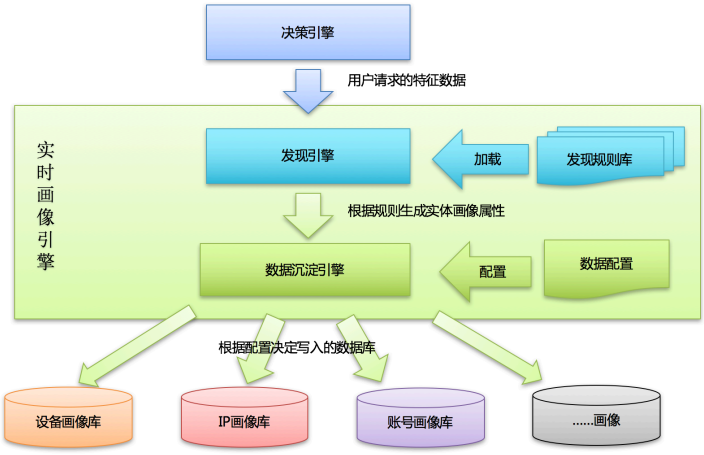

图 27 实时画像架构图

决策引擎的数据给到实时画像引擎的发现模块，发现根据发现规则生成实体画像属性，数据沉淀模块在将数据写入到对应的画像库中。

4.2.8 实时统计引擎技术

实时统计引擎与实时画像引擎技术类似，这不过一个是画像层面的沉淀，一个是基于滑动窗口的实时统计，实时统计可以灵活配置，配置具体业务、滑动窗口、统计指标等参数，对业务请求进行实时统计，在反欺诈往往是为了实时统计出相关异常的指标，提供给实时决策引擎使用，比如 10 秒内发生 100 条相同疑似广告内容，这个统计指标说明这个用户很有可能是黑产广告欺诈用户。

实时统计引擎具体架构如下：

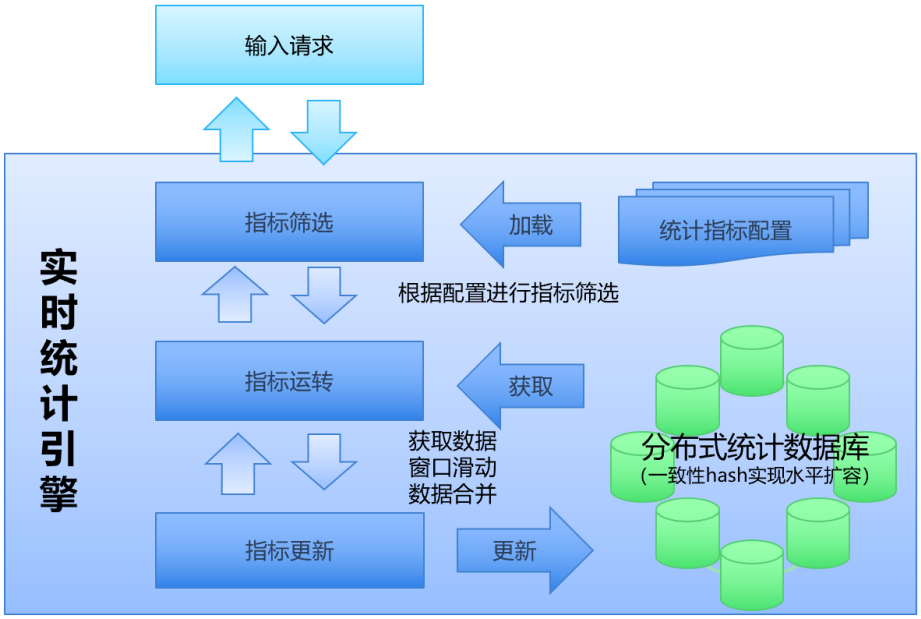

图 28 实时统计引擎示意图

实时统计主要包括五个模块：

- 统计指标配置：配置统计指标，具体包括六个方面，标识、维度、数据、窗口、统计函数、过滤条件；
- 指标筛选模块：加载统计指标配置，并进行指标筛选，检查业

务请求是否存在相关统计指标；

- 指标运转模块：获取当前统计指标相关统计数据，同时进行时间窗口滑动，对窗口的数据进行合并；
- 指标更新模块：将计算的结果更新统计数据库中；

注意，分布式统计数据库：存储统计数据，采用一致性 hash 实现水平扩容。

4.2.9 可信 ID 技术

可信 ID 是基于物理层和协议层信息，结合设备显性标识，生成的移动设备标识 ID，用以标识移动设备的唯一性。可信 ID 不会随设备里的软件变化而变化。IMEI、MAC、设备 SN、蓝牙地址、ICCID 全部可以篡改，唯有可信 ID 是不变的，相当于把每台移动设备实名制了。

可信 ID 有如下关键技术点：

（1）移动应用虚拟执行环境识别：

在通常的情况下，移动应用都运行在一个严格受控的执行环境中，无需关心代码是运行于一台什么样的设备上，是一部真实的移动设备，还是桌面电脑，甚至是浏览器或者其他的什么环境下。可信 ID 的这一虚拟执行环境识别的专利技术，正是要给予应用程序辨识其所处其中的执行环境的能力，检测该环境是否是真正的移动设备环境，从而达到在应用执行环境级别上的真实与虚拟的辨识。

（2）虚拟机/仿真器通用识别：

除了上述的应用执行环境外，更有甚者，其技术可以使用其他的计算机系统来模拟移动设备，甚至是对其进行全方位的仿真。可信 ID 移动设备唯一性甄别技术针对移动设备的硬件制定并实现了特别的探测与鉴定，能够快速、准确地识别出虚拟机、仿真器。

（3）移动设备硬件真伪识别：

除了上述的虚拟与仿真之外，市面上还出现了通过对设备的特定信息进行篡改，试图将同一台设备伪装为无数台不同设备的欺诈手段。可信 ID 移动设备唯一性甄别技术可以对组成一台真实移动设备的关键硬件信息进行采样，并生成唯一编码，即便有部分信息遭到恶意的修改，也能够准确识别出其真实身份。

（4）通过大数据分析移动应用安装意图：

通过对海量用户的数据进行分析发现，专业进行移动应用推广欺诈的从业人员，还会采取批量对真实设备进行自动化操纵来实现虚假安装或者虚假活跃的效果。可信 ID 移动设备唯一性甄别技术同时也对网络、安装行为等数据进行了相应的分析监控，用于监测具有明显刷量意图的渠道或者行为。

4.3 运营商风控技术实践

4.3.1 运营商业务风控系统

运营商风控面向互联网应用的业务风控系统，由黑产知识图谱、风控决策、风险控制等模块组成，可实现风险事前预警、风险的实时

识别、拦截和事后溯源，形成风险的全链路管控。

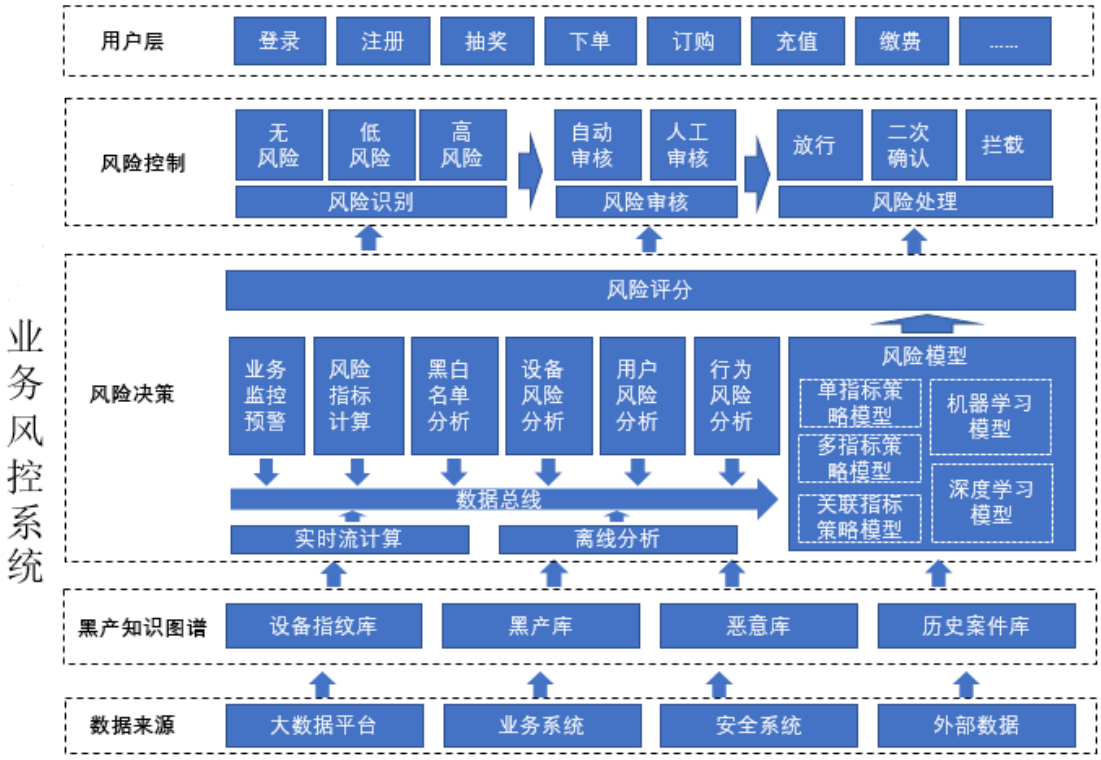

图 29 运营商业务风控系统

● 黑产知识图谱

包含设备指纹库、黑产库、恶意行为库、历史案件库等。

黑产知识图谱的特性如下：

➢ 设备指纹库包含海量手机用户设备指纹数据；

➢ 黑产库包含大量代理 IP 库、大量已沉淀的恶意号码；

➢ 历史案例库包含全国近几年运营商风险案例；

● 风控决策

业务风控系统的核心模块，基于流计算平台和离线计算，策略模型和机器学习模型，对业务进行实时风险分析。

风控决策的特性如下：

➢ 业务处理能力低延时，在较短时间内处理实时请求，用户

无感知。

➢ 支持 Android、iOs、H5、WEB 主流客户端

● 风险控制

包含风险识别、风险审核和风险处理。风险控制可以灵活对不同风险场景、风险点进行灵活配置。

风险控制的特性如下：

➢ 风险识别可以支持不同的评分模型，个性化定制和配置。

➢ 风险处理可以支持滑动验证码、短信挑战码、语音验证码等方式进行二次确认，可以根据不同业务场景进行风险处理配置。

4.3.2 通信数据在风控中的应用

（1）刻画用户通信画像

据工信部数据显示，截止 2019 年 2 月份，我国手机用户总数超 15.6 亿，除去老人和婴儿，手机的覆盖率超 100%。据此通信行为数据已成为个人日常行为的有效载体，分析通信数据、刻画用户通信画像，将成为风控实践的重要手段及标配数据。通过对通信数据的多维度分析，可准确识别用户的行为偏好及消费特征，刻画全面准确的用户画像，为风控环节进行数据储备。

图 30 通信大数据优势

（2）识别逾期风险

逾期与坏账之间具有强关联关系，通过分析通信数据中的通话对象类型（如催收通话、风险号码等），随着时间迁移通话行为的变化情况，可以剖析出逾期用户的通话行为特点，精准识别高风险、高逾期率用户。

（3）社群风险动态防范

团伙欺诈识别是风控环节中尤为重要的一环，分析哪部分人属于同一个微社群及微社群风险评估是识别团伙欺诈的重要手段。社群风险识别透过通信数据图谱中各个节点与其他节点的链接与交互关系，可以分析出远近亲疏以及群体属性结构。并根据社群内各成员的画像信息识别高风险社群，识别团伙欺诈。

（4）协助风险定价及分析偿还能力

通过对通信数据的深入分析，可以精准了解用户的消费习惯，识别用户的消费水平及偿还能力，精准定位用户类型，为金融机构额度授信、风险定价提供数据支撑。4.3.3 通信数据在垂直行业的探索通信大数据在行业的的反欺诈也有重要的应用价值和前景。目前，行业内已探索出基于通信大数据利用机器学习算法进行逻辑加工，针对通信业务消费信息、违约信息、业务状态等上百个通信数据分析，结合金融业（银行系统、消费金融机构）、医疗（社区医疗机构、康养机构、医院）、政务（国家信息中心、保险）等数据堆叠，得到各个自变量的权重，同时计算连续性和类别性变量，得出国际上通行的信用分值结果，从社交圈指标、稳定性指标、履约类指标、安全类指标和消费偏好类指标，构造信誉体系。可以输出：客户标识、用户主手机号、消费情况、流量使用情况、来电稳定性、去电稳定性、蜂巢模型反欺诈等 8 大类共计 88 个细项内容，实现用户通讯行为在金融领域的映射和信息对称，更加直观地为客户画像，有效地甄别目标客户。

4.4 反欺诈效果验证与评估

4.4.1 事前评估

创建反欺诈事前评估流程，在涉及到欺诈行为的业务场景或者营销活动前，先经过反欺诈事前评估通过。具体操作如下：

1、创建某业务反欺诈事前评估工单

2、由风控人员进行反欺诈事前评估

3、反欺诈评估未通过，反欺诈工单处理人进行措施建议，业务侧进行修改。

4、反欺诈评估通过后进行业务活动。

在业务开展前，根据用户风险情报库进行反欺诈评估，使反欺诈评估更快捷。对业务的开展不造成时间上的滞后。

同时，可建立初始用户风险情报库，在营销活动的过程中不断发现风险，总结风险知识，完善用户风险情报库，保持风险情报库的时效性。

每次营销活动后，总结风险，深入分析风险，更新用户风险情报库。

4.4.2 事中分析

对反欺诈行为进行事中风险分析，并对欺诈行为进行相应的管控。将欺诈行为划分风险等级，根据不同的风险等级采取相应的措施。

欺诈风险等级：

- 极高风险；
- 高级风险；
- 中级风险；
- 低级风险；
- 无风险；

在业务过程中，进行事中管控策略，建议以二次校验为主，只对高风险行为进行阻断。管控措施如下所示：

表 6 风险控制与管控策略对应表

风险等级	建议管控策略	风险场景
4 极高风险	2. 拦截阻断	如：在黑名单中、匹配多项高权重策略、机器学习模型评分较高等场景，直接阻断。
3 高风险	2. 拦截阻断	匹配高风险策略，或者匹配多个中风险策略。
2 中风险	1. 二次校验	匹配多个低风险策略，或匹配中风险策略，且完全确定用户行为是恶意行为，采用的管控措施为二次校验。
1 低风险	0. 放行	如：用户通过代理访问系统，只匹配单一低风险策略，采用的管控措施为直接放行。
0 无风险	0. 放行	

在业务开展一段时间后，对事中分析结果进行汇总，并形成报表，用事后评估。

4.4.3 事后评估

对反欺诈行为进行人工验证，采取的验证方式如下:

图 31 反欺诈效果评估体系

为保证用户免收拦截误伤，定期对拦截的反欺诈行为进行效果验证

1、取被拦截的用户进行定期验证。

2、人工进行欺诈行为数据分析验证。

3、判断欺诈拦截行为的准确度。

4、对不准确的欺诈行为拦截，进行反欺诈风险分析模型调整。

5、对于精度不够的反欺诈模型进行精度优化。

欺诈分析需要全方位的分析，包含事前分析，事中分析，事后分析，事后取全量业务数据进行是否欺诈行为分析。比对事中分析中生成的风险报表，评估事中分析与事后分析之间的差距。从以下几个方面进行评估。

- 事中分析的准确度占比
- 事中分析的误伤数据占比
- 事中分析的遗漏反欺诈行为用户占比

根据事后评估，更好的优化反欺诈分析模型，提高反欺诈分析的准确度。

五、移动业务反欺诈的挑战及展望

5.1 反欺诈的困难和挑战

5.1.1 业务风险不确定性分散

相比应用层，业务层具有更大的不确定性和变动性，这会给安全管理带来很大的挑战。频繁的业务变动及难以量化的逻辑会让业务风控在安全管理手段上失控。如，企业面临的海量撞库攻击，在业务繁多的情况下，安全团队可能会跟踪不到被撞的登陆接口 API 的存在。上述不确定性的业务给反欺诈带来了较大的挑战。

5.1.2 风控效果不可判断性高

该问题本质是业务方从数据很难还原出完整的攻击场景。使得决策规则依赖于经验主义，使决策引擎不能持续高效迭代，也让风控处置结果不具有很高的可判断性。例如，安全团队上线了一系列针对恶意注册风控策略，并发现拦截量从 1000 上升到了 5000。从防控的角度看，似乎风控策略起到了一定的防护作用，减少了恶意注册的量，但实际从黑灰产视角看，可能其针对该企业的注册量有数十万，在这个层面上，其实该风控策略是的效果是微乎其微的。

5.1.3 认知盲区不认知性强

黑灰产的快速迭代、精细分工、严密协作，使得其可利用新技术

和资源进行攻击。且攻击手段完全在企业安全团队的认知范围之外。一旦黑灰产攻击的方式超出认知范围，企业的整个风控体系就会失效。如同病毒和疫苗的对抗，已知的病毒已经有了成熟的疫苗可以与之对抗，而一旦新的病毒产生，就需要研制新的疫苗进行防控。

5.1.4 追求数据美观不务实性多

美观的数据背后往往对应难以接受的现实，这造成许多企业不愿正视真实数据。

虚假的移动推广数据，不仅会给企业带来经济损失，还会对企业后续的产品、运营和市场策略都造成方向上的错误，这是个很严重的问题。企业想要越走越远，就有必要掌握移动推广中的真实数据。同样，虚假的移动推广数据也损害了竞争的公平性，以致劣币驱逐良币，不能达到经济可持续发展的效果。只有企业务实，敢于掌握真实推广数据，才能做好反欺诈。

5.2 反欺诈未来展望

5.2.1 加强技术升级优化

面对黑产的技术持续进化，反欺诈技术也需要从多方面、多角度持续迭代更新。

（1）优化设备风险识别技术

移动时代，设备风险识别日益重要，我们需要从设备层面进行攻防研究，做好反欺诈防御的第一道关口，保证设备使用的真实性与有效性，有效识别设备的风险，包括不限于虚拟机、云手机、篡改设备、设备农场、猫池卡池、积分墙等，防止黑产用机器大规模作恶，提高黑产的作恶成本。

（2）优化反欺诈模型

大数据现在已经来临，使用单一维度单一模型，无法兼顾高准率与高召回率的平衡，在高维度海量数据面前，机器学习变得日益重要，黑产在部分领域也已经进化到了深度学习的阶段，反欺诈模型也应由应对的进行更行。要大面积推广机器学习、深度学习，让 AI 为反欺诈技术保驾护航，从海量数据中找到欺诈的线索，挖掘出隐蔽的欺诈团伙，让欺诈份子无地可藏。

（3）优化全栈式实时反欺诈

在移动时代下，反欺诈一定要做到全栈、实时，才能真正的维护好企业安全生态。

一方面，反欺诈不能成为马其洛防线，单点很强，而其他防线薄弱。对于狡猾的黑产，他们寻找的是防守最薄弱的地方，以最低的成本攻破风控系统，实现欺诈获利。，全栈防御应成为重点演进方现，

以保证黑产所有作恶路径上的点都有布防，做到天网恢恢，疏而不漏。

另一方面，反欺诈也不能成为事后诸葛亮，黑产已经获利了结，系统才刚刚发现。黑产是灵活多变的人，作案也是实时变化的，反欺诈系统也必须是实时反欺诈，实时发现黑产的动机，实时拦截黑产的行为，保证系统防御的高效性。

5.2.2 基础共性技术开源

企业在创立初期为了争夺用户和市场份额，往往更加注重产品功能和规模的扩张，风控的建设往往落后于业务发展的需要，给了黑灰产牟利机会。目前整个互联网行业中的风控系统基础设施普及率相对较低，绝大部分企业还处于业务风险对抗的初级阶段。

目前，无论是采购第三方商业风控产品，还是企业自研，对于自身业务在发展中的中小企业，其较大成本都成为了门槛。

未来，底层的风控基础技术应该逐渐开放，使所有企业尤其是中小企业能够快速的拥有基础的反欺诈设施，快速低成本的建立风控体系。再此基础上不断的进行风控策略的优化、风险数据的补充，让企业对抗黑灰产攻击的能力越来越强。

5.2.3 构建产业协作组织

黑产已形成产业联盟，反诈行业同样需要多方共同合作，构建由监管部门、行业协会、金融机构、科技企业共同参与的反欺诈组织，建立数据、技术、人才等方面的合作交流机制，强化同业间风险联防

与合作，提高违约成本。同时，目前移动消费者保护存在很大程度的缺失，维权途径和渠道不畅，因此，要加强行业层面对消费者的权益保护，联合建立客户权益保护中心，建立行业风险缓释与互助机制。

5.2.4 推动完善法制建设

移动时代，技术突飞猛进，业务千姿百态，而法制建设存在一定的滞后性，导致灰色地带产生。黑产利用这些灰色地带肆无忌惮的收割暴利，造成了相关领域的巨大损失。如何界定黑灰产的违法界限，如何界定个人信息保护的边界范围，这些问题相关政策法规的明确。技术与法律在一定情况下存在一定的矛盾性,,前者突飞猛进的发展，后者谨慎的研讨磋商，在这个矛盾空间下，灰色地带总是无法填平。相关部门应加强法律建设,压缩灰产空间,让黑产有更多的法律束缚，让欺诈行为承担更多的惩罚，增加黑产欺诈的犯罪成本。当违法的成本远远大于作恶的收益，黑产自然会选择更有利于他们的一面，放弃欺诈作恶，欺诈的行为也因此减少。

附录 A：移动互联网欺诈模型推演

具体而言，针对一般的移动互联网欺诈，其反应-扩散模型可以写为：（公式标号，加上来源联系）

$$\frac{\partial S}{\partial t} = F(S,I) + D_1 \nabla^2 S$$
$$\frac{\partial I}{\partial t} = G(S,I) + D_2 \nabla^2 I$$

其中 D_1，D_2 是扩散系数。更加具体的来看，可以将上述两个因素更加具体化，假设欺诈人群和欺诈收益都是时间和移动互联网服务市场规模的函数。由于欺诈者数量和欺诈收益之间存在着较强的耦合关系 $F(S,I)$、$G(S,I)$，将其定量化描述为

$$F(S,I) = \frac{\alpha I}{(1+e^{-\mathrm{pI}})sqrt(S)}$$
$$G(S,I) = \frac{\beta A}{(1+e^{-\mathrm{qS}})sqrt(I)}$$

其中 α、β、p、q 皆为对应变量的系数。其中，欺诈者的数量和欺诈收益之间采用 Logistic 方程进行拟合。类似于生物种群和传染病传播等模式，在欺诈的数学模型中，虽然欺诈收益的提升会吸引更多的欺诈者进入该行业，但是欺诈者的增加会使得“竞争”更加激烈，从而降低欺诈带来的收益。

基于实际情况，上述模型的边界条件为：

$$S(0) > 0,\quad I(0) > 0,\quad \left.\frac{\partial S}{\partial t}\right|_{t=0} = 0,\quad \left.\frac{\partial I}{\partial t}\right|_{t=0} = 0$$

附录 B：RETE 算法详解

1） 鉴别网络

由 RETE 算法在进行模式匹配时，是根据生成的鉴别网络来进行的。网络中非根结点的类型有 1-input 结点（也称为 alpha 结点）和 2-input 结点（也称为 beta 结点）两种。1-input 结点组成了 Alpha 网络，2-input 结点组成了 Beta 网络。

每个非根结点都有一个存储区。其中 1-input 结点有 alpha 存储区和一个输入口；2-input 结点有 left 存储区和 right 存储区和左右两个输入口，其中 left 存储区是 beta 存储区，right 存储区是 alpha 存储区。存储区储存的最小单位是工作存储区元素（Working Memory Element，简称 WME），WME 是为事实建立的元素，是用于和非根结点代表的模式进行匹配的元素。Token 是 WME 的列表，包含有多个 WME，（在 Forgy 的论文中，把 Token 看成是 WME 的列表或者单个 WME，为了阐述方便，本文将把 Token 只看成 WME 的列表，该列表可以包含一个 WME 或者多个 WME），用于 2-input 结点的左侧输入。事实可以做为 2-input 结点的右侧输入，也可以做为 1-input 结点的输入。

每个非根结点都代表着产生式左部的一个模式，从根结点到终结点的路径表示产生式的左部。

2） 规则匹配

推理引擎在进行模式匹配时，先对事实进行断言，为每一个事实

建立 WME，然后将 WME 从 RETE 鉴别网络的根结点开始匹配，因为 WME 传递到的结点类型不同采取的算法也不同，下面对 alpha 结点和 beta 结点处理 WME 的不同情况分开讨论。

（1）如果 WME 的类型和根节点的后继结点 TypeNode（alpha 结点的一种）所指定的类型相同，则会将该事实保存在该 TypeNode 结点对应的 alpha 存储区中，该 WME 被传到后继结点继续匹配，否则会放弃该 WME 的后续匹配；

（2）如果 WME 被传递到 alpha 结点，则会检测 WME 是否和该结点对应的模式相匹配，若匹配，则会将该事实保存在该 alpha 结点对应的存储区中，该 WME 被传递到后继结点继续匹配，否则会放弃该 WME 的后续匹配；

（3）如果 WME 被传递到 beta 结点的右端，则会加入到该 beta 结点的 right 存储区，并和 left 存储区中的 Token 进行匹配（匹配动作根据 beta 结点的类型进行，例如：join，projection，selection），匹配成功，则会将该 WME 加入到 Token 中，然后将 Token 传递到下一个结点，否则会放弃该 WME 的后续匹配；

（4）如果 Token 被传递到 beta 结点的左端，则会加入到该 beta 结点的 left 存储区，并和 right 存储区中的 WME 进行匹配（匹配动作根据 beta 结点的类型进行，例如：join，projection，selection），匹配成功，则该 Token 会封装匹配到的 WME 形成新的 Token，传递到下一个结点，否则会放弃该 Token 的后续匹配；

（5）如果 WME 被传递到 beta 结点的左端，将 WME 封装成仅有一个 WME 元素的 WME 列表做为 Token，然后按照（4）所示的方法进行匹配；

（6）如果 Token 传递到终结点，则和该根结点对应的规则被激活，建立相应的 Activation，并存储到 Agenda 当中，等待激发。

（7)如果 WME 被传递到终结点,将 WME 封装成仅有一个 WME 元素的 WME 列表做为 Token，然后按照（6）所示的方法进行匹配；

以上是 RETE 算法对于不同的结点，来进行 WME 或者 token 和结点对应模式的匹配的过程。