

2021 网络金融黑产研究报告

中国工商银行金融科技研究院

2022 年 02 月

版权声明

本文版权属于中国工商银行所有，并受法律保护。转载、摘编或利用其他方式使用本文观点或文字的，应注明来源。

违反上述声明者，将追究其相关法律责任。

摘 要

自网络金融诞生以来，黑产一直是各网络金融企业面临的重要威胁，随着黑产的发展与扩张，各类新型的欺诈场景、欺诈技术不断渗透到针对金融企业的攻击中，其链条化分工、跨行业作案等特点，为金融机构的风险防控带来了更大挑战。为了在黑产防护工作中做到有的放矢，我部组织金融科技研究院安全攻防实验室对 2021 年网络金融黑产进行了跟踪分析，从技术手段特点、欺诈手法等方面对黑产 2021 年的活动趋势与特点进行了剖析，并提出了黑产防控工作的思路。2021 年网络金融黑产技术手段不断升级，钓鱼、人脸、跑分、引流、退保五个关键词概括了黑产一年的发展趋势。针对黑产的新变化，报告中从规划、合作、样本、AI、情报等 5 个维度给出了进一步拓展银行业的安全防护思路。

目录

一、 序言.....	1
二、 2021 网络金融黑产发展趋势.....	2
(一) 关键词一 钓鱼：技术、场景双升级，钓鱼攻击重回欺诈兵器谱榜首.....	3
1、 自动变人工，骗子实时值守钓鱼.....	3
2、 钓鱼“破圈”，基于企业内网邮件的钓鱼需要进一步关注.....	4
3、 打破“次元壁”，黑产发动三次元钓鱼.....	5
(二) 关键词二 人脸：安全不能只靠“颜值”，人脸识别面临新挑战.....	6
1、 黑客自制手机系统，人脸难逃劫持命运.....	7
2、 视频来电慎接听，人脸窃取需防范.....	8
3、 合成人脸瞒天过海，刷脸也有“万能钥匙”.....	9
(三) 关键词三 跑分：跑分洗钱，黑产团队各显神通.....	9
1、 “收款码跑分”，收款二维码或成黑产洗钱新途径.....	10
2、 跑分参与人员泛化，业余兼职竟成洗钱帮凶.....	10
(四) 关键词四 引流：黑产引流，社交 APP 成不法行为推广新阵地.....	11
1、 “震惊体”泛滥朋友圈，点开你就中招了.....	11
(五) 关键词五 退保：代理退保，保险业与黑产近身遭遇战.....	13
1、 代理退保：为消费者设下的收割“连环套”.....	13
2、 代理退保向链条化发展.....	14
三、 2021 黑产对抗思考.....	15
(一) 维度一 规划：建设由业务驱动的企业安全架构.....	15
(二) 维度二 合作：跨机构、跨行业的黑产联防联控.....	17
(三) 维度三 样本：小样本掘金，助力反欺诈精度提升.....	18
(四) 维度四 AI：AI 安全对抗平台建设.....	19
(五) 维度五 情报：知己知彼，黑产情报监控分析体系探索.....	20
四、 结束语.....	21

一、序言

随着新一轮金融科技的发展与产业的变革，金融行业加快了数字化转型的步伐，人工智能、云计算、物联网等技术与金融业务的进一步融合成为推动金融创新转型的新引擎，为服务实体经济、化解金融风险提供了新的途径与抓手。灵活与便捷的金融业务模式在为用户提供了更加优质的金融服务的同时，也面临着来自黑产不断演变迭代的各类新型攻击威胁。

自网络金融诞生以来，黑产一直是各网络金融企业面临的重要威胁，随着黑产的发展与扩张，各类新型的欺诈场景、欺诈技术不断渗透到针对金融企业的攻击中，其链条化分工、跨行业作案等特点，为金融机构的风险防控带来了更大挑战。



图 1 2021 年黑产趋势热词

中国工商银行金融科技研究院安全攻防实验室（以下简称“实验室”），作为国内金融业首家体系化黑产研究机构，多年来持续面向社会发布黑产动向及防护技术的研究报告，近年来成功预测了包含人脸识别欺诈、屏幕共享泄密等多种新型黑产趋势，为金融行业的黑产防护提供了探索思路。随着 2021 年的结束，实验室继续从金融安全从业者的角度，为大家带来 2021 年黑产的新趋势与新变化，希望能够为全社会带来黑产防护的新思路。

二、2021 网络金融黑产发展趋势

疫情以来，常态化防控机制进一步推进业务线上化，各类无接触办理、云网点等新型业务模式蓬勃发展，使得网络金融黑产防护整体趋势愈加严峻。移动互联网、人工智能技术的快速发展使得黑产作恶手段进一步变化，众包作案、跑分洗钱等新型欺诈手法进一步扩大了黑产的参与人员数量及影响面，让各类传统的黑产防护手段面临更大挑战。

针对黑产发展的严峻趋势，公安机关通过断卡行动、国家反诈 APP 等一系列手段对黑产进行打击，有效降低了黑卡数量，阻断了黑产活动的源头。在国家的重拳打击下，黑产的活动频率及规模已经连续数月呈下降态势，结合年内反电信网络诈骗法草案的提审，对黑产从业人员起到了有效的威慑，黑产进一步向隐蔽化、随机组团化转移。

为了巩固年内的黑产防护成果，进一步压降黑产的活动规模，我们对年内出现的各种新型网络金融黑产欺诈手法进行了研究汇总，用

于提升全社会对黑产形势的了解，做到知己知彼，我们总结的 2021 年黑产趋势的五大关键词分别是“钓鱼”、“人脸”、“跑分”、“引流”、“退保”。

（一）关键词一 钓鱼：技术、场景双升级，钓鱼攻击重回欺诈兵器谱榜首

钓鱼攻击作为最早的黑产欺诈手法之一，在国内出现时间已经超过了 20 年，在很长一段时间中都是占比最高的欺诈手法，这也使得黑产习惯性的将欺诈受害人称呼为“鱼”。近年来，随着金融机构和用户的防护手段及安全意识的提升，使得传统成本低、风险小、广撒网、多敛鱼的钓鱼模式在前几年已经不再具有收益优势。今年我们接触的各类新型钓鱼趋势中，体现了黑产从不同维度对钓鱼攻击手法的升级，钓鱼类攻击的数量同比去年增长了 33%，钓鱼攻击再度成为占比最高的欺诈手法。

1、自动变人工，骗子实时值守钓鱼

传统的钓鱼欺诈中，往往通过广撒网、多敛鱼的模式大规模推送钓鱼信息，后台通过脚本自动化地针对钓鱼获取的账户进行资金窃取，但是由于基于脚本自动执行，往往成功率较低，作案金额也较小。在今年钓鱼后台服务器的溯源反制中，实验室发现基于人工进行“实时值守”的钓鱼占比逐渐提高。与以往最常见的自动化钓鱼后台不同，实时值守钓鱼中，不法分子守在钓鱼后台页面等待受害人“上钩”，

当有受害人通过钓鱼页面提交了个人敏感信息后，钓鱼者会根据受害人账户的情况，向其反馈不同的内容，比如为了欺诈金额最大化，会向受害人反馈“请反馈账户余额”，如受害人资金较少，会向其反馈“请向账户内转入 5000”等页面内容，如果受害人提供的密码、验证码不正确，还会通过页面二次向受害人询问。

ID	开户行	姓名	卡号	身份证号	手机号	动作	IP	系统	CVV
104	储蓄卡	银行				未通过		苹果	
103	储蓄卡	银行				未通过		苹果	
102	储蓄卡	银行				【通过】 请向账户内转入 5000 元以验证账户余额		苹果	
101	储蓄卡	银行				预留手机号错误，单独输入手机号		安卓	
100	储蓄卡	银行				不支持该银行卡，请更换其他银行卡号进行认证！		安卓	
099	储蓄卡	银行				不支持【信用卡】，请更换【储蓄卡】		苹果	
098	储蓄卡	银行				【通过】 请向账户内转入 5000 元以验证账户余额		苹果	
097	储蓄卡	银行				您输入的信息与银行记录的信息不一致		苹果	
096	储蓄卡	银行				请输入正确密码		苹果	
095	储蓄卡	银行				CVV 错误，请重新输入有效期和 CVV		苹果	
094	储蓄卡	银行				【证件】错误，单独输入身份证		苹果	
093	储蓄卡	银行				【卡号】错误，单独输入银行卡号		苹果	
092	储蓄卡	银行				【姓名】错误，单独输入姓名		苹果	
091	储蓄卡	银行				【通过】 审核通过==24小时内审核完毕		苹果	
090	储蓄卡	银行				请输入网银密码		苹果	
089	储蓄卡	银行				【支付宝】账号+登录密码+支付密码		苹果	
088	储蓄卡	银行				【不支持储蓄】更换名下信用卡		苹果	
087	储蓄卡	银行				当前卡余额不足以认证核实！请存入 5000 元后再尝试		苹果	
086	储蓄卡	银行				当前系统繁忙，请 5 分钟后尝试认证		苹果	

图 2 钓鱼服务器后台样本示例

2、钓鱼“破圈”，基于企业内网邮件的钓鱼需要进一步关注

决定欺诈成功与否最重要的是剧本的选择，近年来黑产钓鱼的剧本持续向着精准化、定制化方向迭代，主要围绕用户个人生活场景，今年黑产中出现了由基于生活圈向工作圈渗透的破圈趋势。



图 3 基于企业内网邮件体系的钓鱼邮件样本

黑客通过网络入侵等手段，利用企业邮件系统，仿冒人力、财务等部门，向员工的企业办公邮箱发送钓鱼邮件，一般通过“补贴发放”、“工资补发”等具有吸引力的标题进行诱骗。受害人点击后，再以“个税补缴”、“发放补贴”等名义诱导用户填入敏感信息。

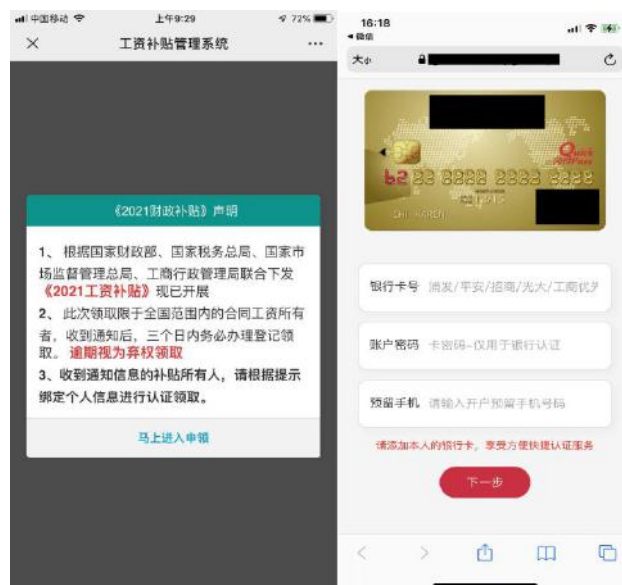


图 4 企业邮箱钓鱼页面

由于邮件来自企业办公邮箱，此类场景下受害人一般缺少反欺诈意识，欺诈成功率明显增高，此类跨界钓鱼的欺诈手法在今年有明显的增长趋势。

3、打破“次元壁”，黑产发动三次元钓鱼

钓鱼攻击作为互联网兴起后所产生的黑产作案方式，一直以来主要停留在网络上，今年，“立体式”钓鱼引起了广泛关注。

今年 11 月，公安机关破获了一起利用实体虚假 ATM 进行钓鱼攻击的案件。案件中，不法分子从网上采购了一套 ATM 设备，在租赁的

临街店面中对外部署，伪装成某银行 ATM。



图 5 虚假钓鱼 ATM

当受害人来进行取款业务时，仿冒 ATM 会对前来取钱用户的银行卡磁道、密码等信息进行记录，随后制作复制的银行卡后前往外省或境外进行盗刷。

10 年前国内也曾出现过假冒 ATM 的事件，但是由于山寨 ATM 过于粗制滥造，根本不会造成危害，而本次事件中，钓鱼使用了真实 ATM、近乎真实的门脸部署模式，使得一般用户无法辨别真伪。由此可见，黑产的钓鱼攻击已不仅仅局限在传统的网络空间中，线下立体空间的钓鱼风险也是重要的关注点。

（二）关键词二 人脸：安全不能只靠“颜值”，人脸识别面临新挑战

目前，疫情防控常态化使得无接触认证、无感认证等身份识别需求激增，人脸识别技术凭借其便捷易用的特点，在互联网金融领域发挥着重要作用。随着人脸识别攻防技术不断发展和反复博弈，人脸识别应用场景也成为了黑产对抗的主战场，除大家熟知的照片活化攻击

外，今年又出现了新的攻击手法，黑产精心设计了人脸欺诈场景并使用了新型攻击技术，让“刷脸”再次面临新挑战。

1、黑客自制手机系统，人脸难逃劫持命运

安全行业与黑产在人脸识别领域的攻防博弈已经经历过多次迭代，从早期的图像级到中期的应用级，再到后期的算法级，金融行业和安全公司通过一系列手段对黑产的攻击方式进行了有效防护，然而对抗并未停止，人脸攻防进一步深入到了移动操作系统，为企业防护带来了新的挑战。黑产制作了针对人脸识别进行攻击的定制 ROM 专用操作系统，实现绕过人脸识别。



图 6 黑产定制 ROM 工具示例

通过对黑产情报渠道的监控发现，目前黑产从业者已将不同机型与定制 ROM 进行了整合，并通过黑产渠道贩卖，帮助下游黑产绕过人脸。

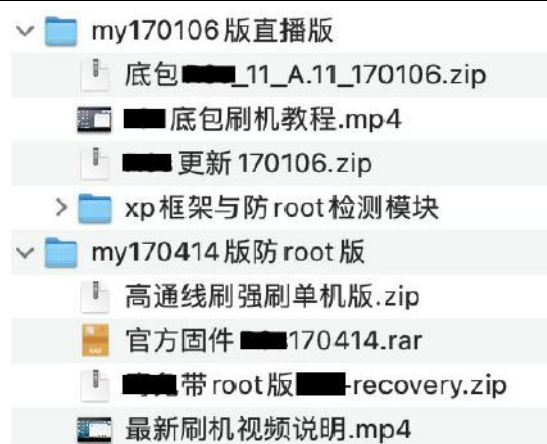


图 7 网上售卖的 ROM 注入工具示例

为解决此类风险，需要执法机构加大对非法工具售卖的打击与治理，从源头上封杀此类技术的传播。

2、视频来电慎接听，人脸窃取需防范

人脸隐私泄露一直是社会关注的热点话题，随着相关法律法规的不断完善，人脸数据的使用和存储得到了规范，增大了黑产获取人脸信息的难度。黑产从业者精心设计了新的欺诈场景，通过视频通话窃取受害人的的人脸信息。黑产从业者一般假冒公检法机关或银行工作人员，恐吓受害者涉嫌“洗钱”、“藏毒”、“欠贷”等案件，要求受害者通过视频通话进行身份核实。视频通话过程中，黑产从业者要求受害人完成指定人脸动作，同时开启录屏功能获取受害人的的人脸信息。

与利用活化软件生成的合成视频相比，通过视频通话获取到的人脸视频是受害者本人完成的人脸识别动作，不存在合成及伪造特征，很难被人脸算法识别。考虑到人脸特征具有唯一性，一旦被黑产窃取，将直接导致人脸认证失效，造成资金和财产损失。因此在接听视频来

电前，一定要核实对方身份信息。

3、合成人脸瞒天过海，刷脸也有“万能钥匙”

攻防实验室近年来持续跟踪新型生物识别攻击技术，监测发现一种新型攻击方法需要安全从业者加以足够的重视。今年，以色列研究人员发现一种创建“万能人脸”的方法，与“万能钥匙”功能类似，使用一张“万能人脸”即可通过多人的人脸身份认证。研究人员基于 StyleGAN 图像生成算法合成候选人脸，再利用有限内存矩阵自适应进化策略筛选出最佳的“万能人脸”。

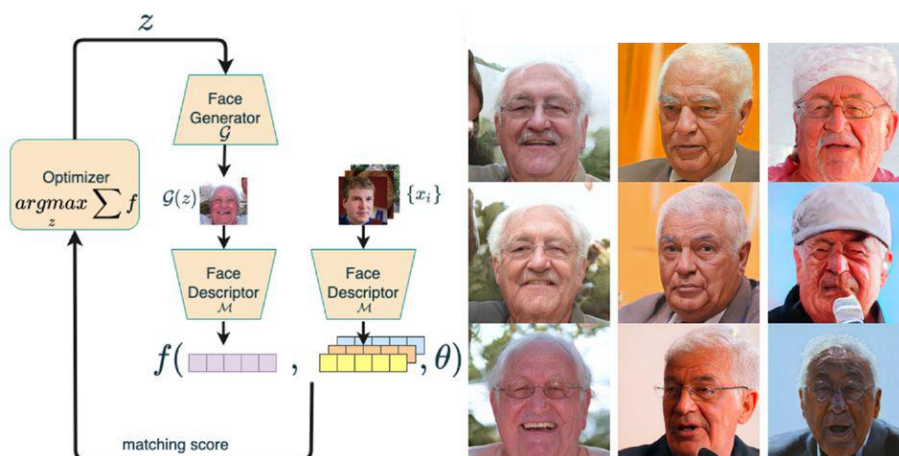


图 8 万能人脸示技术及效果图¹

上图中的 9 张人脸即为算法合成的万能人脸，视觉上看这些人脸与普通照片并无二致，但是却可实现与任意人脸的匹配，识别通过率最高达到 64%左右。

（三）关键词三 跑分：跑分洗钱，黑产团队各显神通

随着人工智能、大数据分析等新技术在反欺诈领域的运用与推

¹图片摘自论文《Generating Master Faces for Dictionary Attacks with a Network-Assisted Latent Space Evolution》

广，传统的黑产洗钱手段已经被风控模型有效遏制。为了规避风控模型，一种名为“跑分”的新型洗钱手法逐渐在黑产团队间蔓延开来。

1、“收款码跑分”，收款二维码或成黑产洗钱新途径

所谓“跑分”，就是指利用第三方支付账户，帮助不法分子转移“黑钱”。这种新型黑产洗钱途径具体为，跑分参与者向跑分中介平台缴纳保证金，并将自己的收款码提供至平台，黑产从业者需要洗钱时，会向跑分人员的收款码进行转账，每当跑分人员收到一笔款项时，其平台中的保证金就会相应减少，由此来完成一进一出的“跑分”行为。参与者完成相关资金转移后即可获取对应佣金。

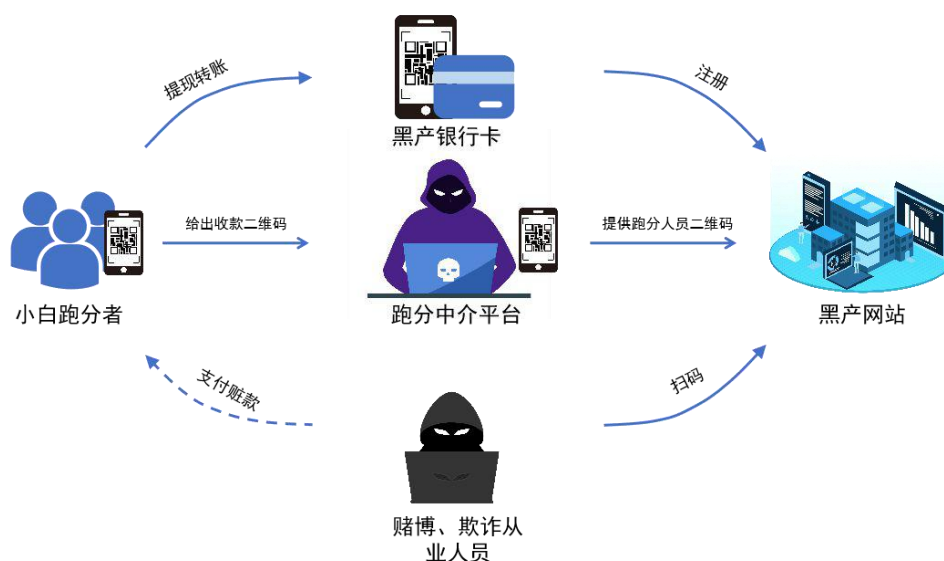


图9 二维码跑分洗钱流程图

2、跑分参与人员泛化，业余兼职竟成洗钱帮凶

传统洗钱的核心工作落在黑产团队内部的“卡农”身上，“卡农”一般持有多张实名且非本人注册的银行卡，通过这些银行卡转入转

出，完成洗钱工作。但随着监管力度进一步收紧，黑产的洗钱手段也进一步隐蔽化。为了逃避风控模型的打击，黑产团队将跑分洗钱活动披上业余兼职的外衣，包装为兼职工作。因此，在跑分活动中包含了大量非职业黑产人员的参与，将洗钱风险转向普通民众，以“学生兼职”、“农民工兼职”、“高薪日结”等标题，吸引普通民众参与跑分，由于跑分人员使用的跑分账户往往是日常生活中使用的银行卡，其中包含了工资、缴费、消费等大量正常的用户交易记录，使得模型防控的难度进一步增大。

（四）关键词四 引流：黑产引流，社交 APP 成不法行为推广新阵地

当前互联网渠道发展越发迅猛，用户流量成为重要资源，黑产对用户的引诱方式向着更隐蔽、更难以识别的方向发展。前期实验室曾对黑产 SEO（搜索引擎作弊）进行过专项研究，其中涉及的“黑产引流”推广手法在今年明显增多，由于黑产引流的隐藏手段花样百出，使得其识别难度与危害也进一步加深。新华社、人民网、光明日报等多家媒体都对此新型黑产引流手法进行了专项报道。

1、“震惊体”泛滥朋友圈，点开你就中招了

《震惊！国人看完都转了》、《重大违纪来袭，内幕曝光》，这些成为公众号、朋友圈中越来越频繁出现的标题内容，背后却往往隐藏着利益收割。黑产利用夸张的标题、图片生成相关的文章，当用户

被引诱点击进入此类标题的文章后，会被黑客通过恶意代码、窗口广告、甚至是系统漏洞引导至黄赌毒等非法网站，从而获得非法网站的广告提成。我们对“震惊体”网页背后涉及的恶意代码及业务关系进行了跟踪，此类网站使用了大量代码混淆及隐匿技术，经过梳理分析后，可以看到其背后错综复杂的关系网。

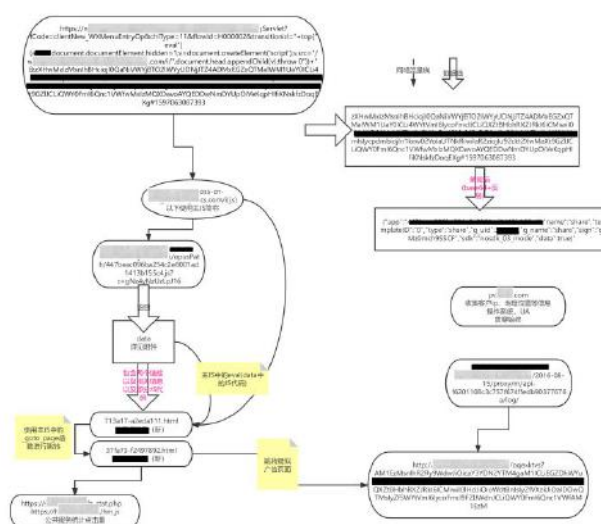


图 10 黑产引流网站流程图

此类“震惊体”引流手法相较于传统黑产传播存在以下三个特点：

扩散性强——黑产利用诱骗或恐吓的标题吸引用户点击链接，并将内容分享至群聊或朋友圈。普通用户难以分辨真假，极易成为黑产推广的传播工具，使得黑产内容成病毒式扩散。

封禁难度高——恶意链接借助各大知名网站的子域名，构造可重定向的 URL 向用户恶意推广。由于知名官方域名下有大量合法运营的业务，社交 APP 无法直接封禁域名，从而有效增加了恶意链接的存活时间。

易构造变形——引流页面遭到查封时，黑产可基于页面模板快速伪造新的引流页面，将页面内容修改为热点内容即可重新发布新的链

接，诱骗用户点击。

今年公安机关打掉了多个职业从事“震惊体”网文制作的黑产团伙，但由于此类引流方式的特征隐蔽、作案成本低，使得此类手法的打击难度较大。

（五）关键词五 退保：代理退保，保险业与黑产近身遭遇战

今年，名为“代理退保”形式的恶意代理投诉案例在保险行业呈快速增长态势，黑产通过社交平台、短视频平台等渠道，煽动消费者委托其代理“全额退保”，并收取高额费用，一些地区甚至出现了专门以“代理全额退保”、“代理投诉”为业务的组织和机构，一条围绕“代理退保”等形式的保险业黑色产业链已然形成，扰乱保险市场环境，侵害保险消费者的合法权益，针对此类行为，最高法已经出台了司法解释，各地均针对此类行为进行了专项打击。

1、代理退保：为消费者设下的收割“连环套”

“代理退保”是一种披着维权外衣的非法牟利行为，黑产以可帮助客户全额退保为名，引诱贪小便宜的保险客户申请退保。过程中黑产会收集客户的保单、身份证、银行账号等隐私信息，随后指导客户套取保险销售员违规行为，甚至直接伪造违规情节、炮制“证据”，向监管部门进行恶意投诉。当客户后获得全额退保后，黑产会从中赚取高达 60% 的高额服务费，部分黑产甚至一石多鸟，将客户隐私信息

转卖给下游黑产，对客户进行二次诈骗，将“客户价值最大化”。

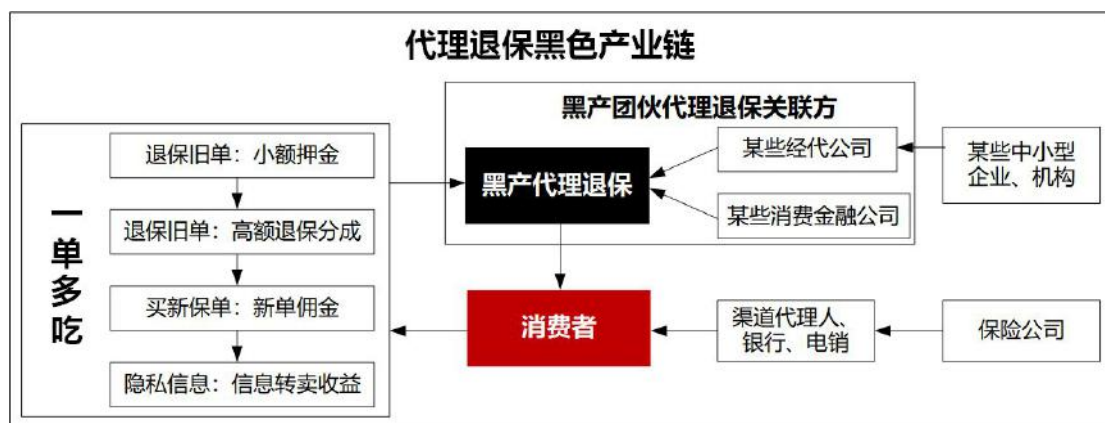


图 11 一单多吃:代理退保欺诈流程

2、代理退保向链条化发展

随着不法分子在代理退保中尝到甜头，此类作案手法逐渐向链条化发展。黑产团伙公开招揽退保推广人员，在社交媒体平台开班授课，培养“下线”团队，甚至在电商平台开立“保险维权”、“退保咨询”等店铺，各类“全额退保”教程层出不穷，由于此类黑产行为特征隐蔽、违法边界模糊，使得其规模越发扩大，甚至出现产业化的特点。目前经过各电商平台、社交媒体的打击治理，此类信息已经有效收敛。



图 12 电商中代理退保教程售卖

三、2021 黑产对抗思考

“钓鱼”、“人脸”、“跑分”、“引流”、“退保”五个关键词概括了黑产一年来的作案趋势。作为安全从业者，现有的黑产防护已经不能停留在“头疼医头、脚疼医脚”，需要从更高维度视角给出防护思路，做到“快敌一步”。为此，我们提出了五个维度的应对思路，从根源提升安全防护，实现对黑产防护的新思考与新探索。五个维度的关键词分别是：“规划”、“合作”、“样本”、“AI”、“情报”。

（一）维度一 规划：建设由业务驱动的企业安全架构

欺诈黑产攻击主要由业务安全风险问题引发，因此金融机构在进行企业安全架构规划时，不仅需要考虑网络安全风险，也要重视业务安全风险的系统规划。但目前大部分企业进行企业安全架构规划时，往往更加重视网络安全规划，体系化的业务安全受到的关注较少。这种规划方式容易导致出现“打补丁式”的黑产应对模式。我们可以形象的把这种情况比喻为“冰山效应”：即只考虑“冰山上”的部分，而较少考虑“冰山下”的隐藏部件，这些隐藏部件与业务需求直接相关，如域关系、信任关系、风险评估等，会直接影响业务的认证措施、风控策略等能否有效涵盖业务安全风险。

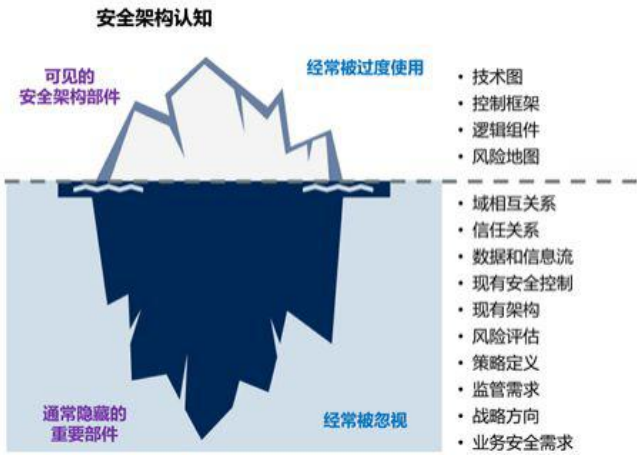


图 13 安全架构“冰山”²

安全架构是企业架构的一部分，安全架构与企业架构密切结合，对此，SABSA、TOGAF、Gartner 等机构提出了一系列由业务需求驱动的企业安全架构方法论。其中，舍伍德（SABSA）是业界较为推崇的一套方法论，它提出了一种以业务需求为基础构建的五层企业安全架构，由业务结果驱动企业架构建设，以确保安全架构基于业务风险的部署具有针对性，并满足业务目标。

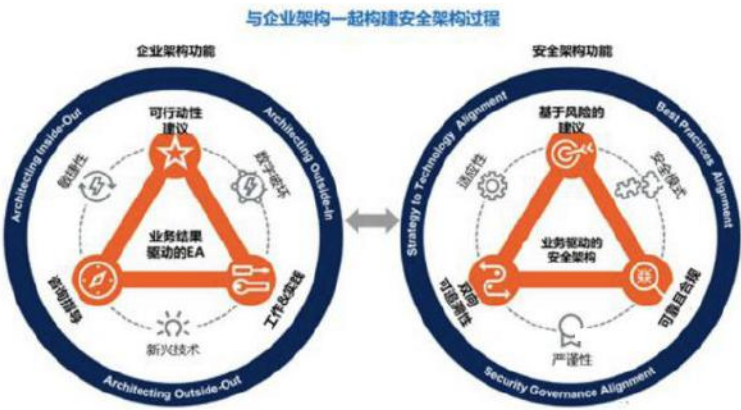


图 14 业务架构与安全架构关系图³

工商银行目前已建立较为完善的业务驱动的企业架构方法论，在此基础上，实验室经过对舍伍德方法论的探索，依托工行业务架构，

²图片来自 Gartner
³图片来自 Gartner

开展了以业务架构驱动的安全架构建模模型的探索研究，可以为后续国内大型金融企业开展业务视角的企业安全架构建设提供有益建设思路。

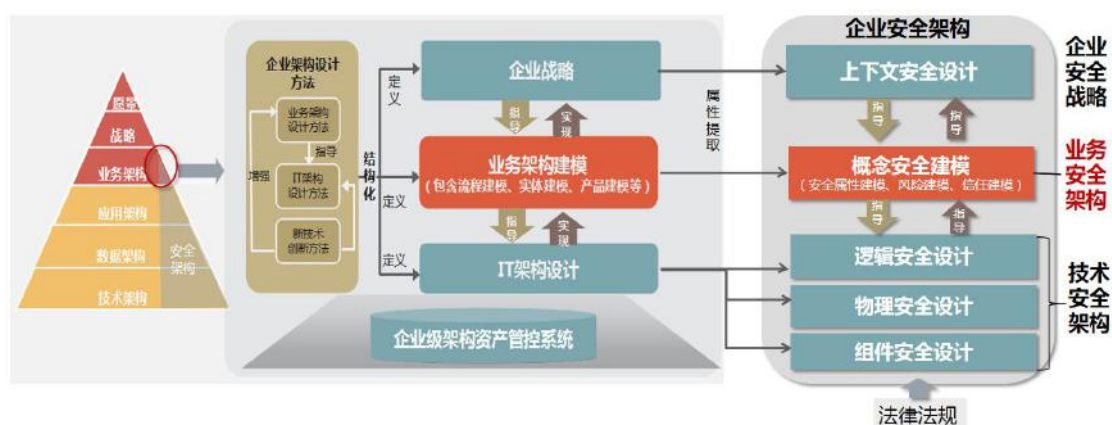


图 15 工行企业安全架构

（二）维度二 合作：跨机构、跨行业的黑产联防联控

跨行业、跨机构的金融数据共享问题已经成为行业发展主要痛点和诉求，也是风险管控跨行难的主要原因。当前不同金融机构间账户资金往来极其便捷，一笔资金可以在短短几分钟内流转至多家金融机构，同一客户也可能在多家银行申请信用卡或贷款，频繁的资金往来为识别资金的真实来源和流向、规避恶意信用透支、洗钱等风险带来很大挑战。而各家金融机构目前的运行机制普遍存在三个问题，即“不敢共享”、“不愿共享”、“不便共享”，很难达到“风险联防联控”的要求。

上述三类问题需要探索突破跨行数据共享难题的解决方案。金融机构可以通过推动客户金融数据跨行的合理适度应用和安全共享，加强客户风险防控，提升跨行联动的风险防控能力。一是借助跨行数据

反映出客户是否有异常诈骗行为，在监管指导下开展风险账户信息共享，跨行联动做好反欺诈风险防控；二是可以通过跨行金融数据对客户资信情况进行核查；三是可以借助跨行数据透彻了解贷款资金流向，加强贷后监测及预警能力。

目前，国家已经从政策层面鼓励合法的数据交易，北京、深圳、上海等地已经开展了数据交易所相关的筹备、建设工作，将在跨机构、跨行业的数据共享与黑产防控方面发挥重要作用。

（三）维度三 样本：小样本掘金，助力反欺诈精度提升

专家规则作为传统的反欺诈手段，存在更新不及时、误报率高、维护费用昂贵等瓶颈，不能准确高效地对跑分、真人众包等新型黑产攻击进行欺诈判别，而在将人工智能模型应用于反欺诈的过程中，又往往面临缺少足够欺诈样本的问题。为缓解样本缺乏的痛点，小样本学习作为新兴的前沿研究领域，已成为业界研究热点。

学术界在小样本学习领域的解决方案主要有两个方向，一是利用无监督异常检测算法挖掘异常数据进行重点剖析，比较异常数据与正常数据的差异性，总结提炼异常样本特征；二是通过迁移学习、贝叶斯方法、强化学习等手段解决样本标签不足的问题。在该领域，多伦多大学的研究人员提出了一种经典的基于原型网络（Prototypical Networks）的小样本学习方法，其主要思想是将每个输入样本映射到高维特征空间中，计算高维特征向量的均值作为原型中心，并基于欧几里得距离度量各样本到各个原型中心的距离，实现样本类别预测。

后续如何将前沿学术研究成果应用于金融反欺诈领域，仍需不断探索与实践。

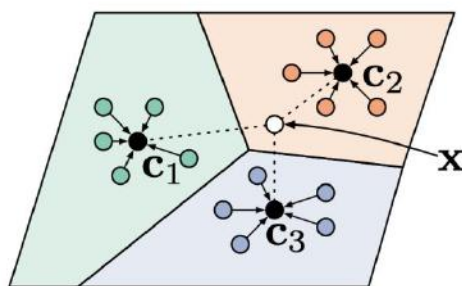


图 16 小样本学习原型网络图⁴

上图展示了小样本场景在二维空间下的原型网络， C_1 、 C_2 、 C_3 是通过计算均值得到的每个类别的原型中心，可根据未知样本 x 到原型中心的距离对 x 进行分类。

（四）维度四 AI：AI 安全对抗平台建设

从近期黑产各类新型攻击手法及趋势可以看出，人工智能技术在黑产攻防中扮演着越来越重要的角色，从中衍生出的安全问题已经成为制约其深度应用的重要因素，国内外高校和企业均成立了多个人工智能安全实验室，保障人工智能的安全发展。从国内外的研究现状和技术发展来看，多个人工智能技术头部企业已在 AI 安全检测领域开展积极探索，其中微软研发了人工智能安全风险评估工具 Counterfit，基于预制的攻击脚本集，对图像识别、文本分析等模型开展自动化渗透测试，实现对逃逸攻击、模型窃取等漏洞的针对性检测。

⁴图片摘自论文 Prototypical Networks for Few-shot Learning

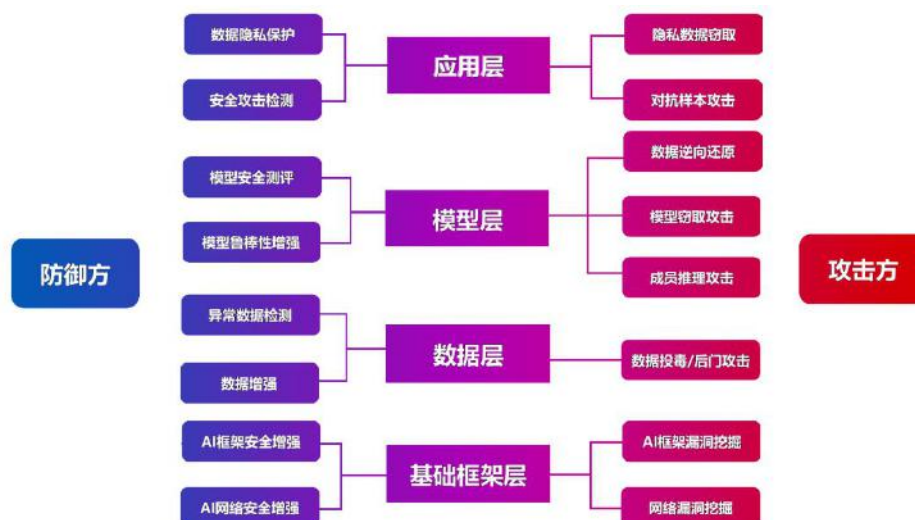


图 17：人工智能对抗平台

目前业界暂未形成成熟统一的人工智能安全评测方案，人工智能安全不能简单的一事一议，未来技术发展依然任重道远，需要业界共同开展人工智能安全测评技术研究，通过构建人工智能安全对抗平台，针对生物识别、智能客服等人工智能细分技术领域，形成基于不同领域的人工智能系统安全性检测与评估标准体系，如持续开展人脸识别攻防对抗，基于查询攻击、迁移攻击等对抗攻击方式对人脸识别算法进行安全检测，辅助提升算法的识别精度和鲁棒性，抵御合成人脸、万能人脸等攻击手法，实现系统化提升人工智能安全可信能力。

（五）维度五 情报：知己知彼，黑产情报监控分析体系探索

随着国家不断加大黑产打击力度，黑产对抗已成为常态化、日常工作。黑产上下游产业链协同作案的特点使得攻击面及攻击强度进一步提升，对防护和溯源提出了新的挑战。黑产间协同作案普遍依赖群聊、论坛、暗网等渠道进行信息沟通和工具买卖。因此，通过覆盖网络全域的爬虫技术及人工智能语义分析技术对黑产协作交流渠道

进行监控，能够帮助金融机构及时获取黑产攻击动向，进而实现针对性反制。

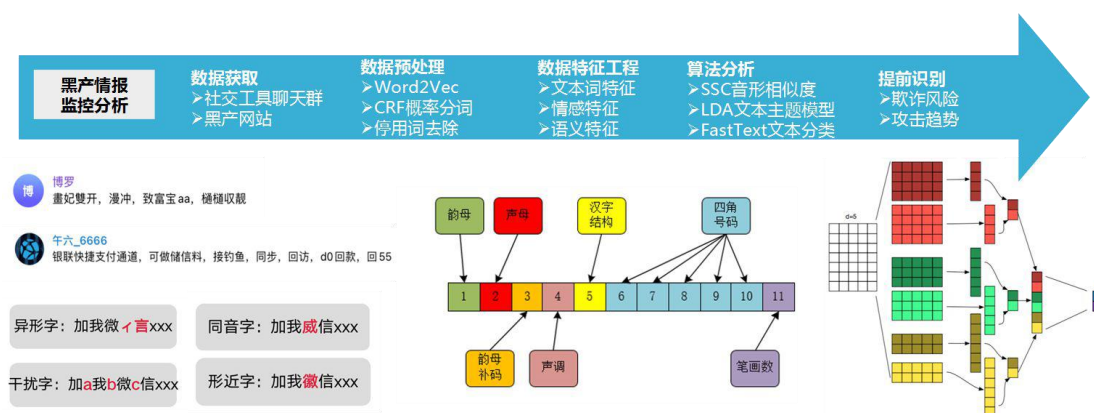


图 18 黑产情报分析技术示例

在实际监控过程中发现，黑产为了隐藏攻击信息采取了一系列手段，一方面利用暗网、地下沟通工具等技术来掩盖行踪，另一方面使用大量异型字、同音字、专业黑话等避免检索封禁，这对情报获取和监控提出了新的挑战。为深入分析黑产情报，需要结合自然语言处理、知识图谱等技术，关联多源碎片信息，对变形和转义的黑产信息进行还原，基于快速文本分类网络（FastText）和文本主题模型（LDA）实现对黑产最新动向的实时感知和对黑产攻击的精准预警与快速响应，提升黑产防护能力。

四、结束语

在产业数字互联的大时代中，黑产防护与数字化是一体两面，随着业务边界的扩大与增长，黑产隐蔽化、技术化、产业化的特点也愈发凸显，单靠某个企业单打独斗、闭门造车必然无法应对未来千变万化的黑产形式。需要通过国家、行业、机构多个层面的联合共建，强

化黑产的抵御之路。

目前，工商银行正积极布局黑产防护生态建设，结合自身长期积累的安全经验，践行大行担当，与金融同业携手筑牢数字化转型的安全基石，做好客户身边“可信赖的银行”。