登录

【反欺诈专栏】关于IP，这里有你想知道的一切！下篇

同小盾 / 2017-07-13 03:29:00 / 浏览数 4789 技术文章 技术文章 顶(0) 踩(0)

IP地址的所有研究，其实可以归到三个问题上：

1、这个IP在哪儿?
2、这个IP是什么?
3、这个IP干了什么?

本篇是关于IP地址研究的最后一篇文章，我们来聊一聊：如何分析一个IP地址做了什么?

上一篇中，我们提到了通过每个IP下的用户行为，判断用户群体的类型，进而给出IP地址的分类结果。那些，其实都只是最简单，最基础的一些分类方法。通过特定的指标计

然而，同盾的IP画像，是围绕反欺诈展开的，我们希望能够准确的评估一个IP地址的风险性，进而在风控策略中进行调控。

在IP画像设计初期，我们设计了一个风险评分，用于总体评价这个IP地址风险。风险分数中，IP是否有代理行为、是否命中已知的威胁情报、是否发生过风险行为，都作为评

比如，我们曾经发现过一个IP地址，由于频繁的发生盗卡行为，最终我们给出的风险评分达到了94分(0~100，分数越高风险越高)，然而这个IP下其他行为都是正常的，大量

于是，我们萌生了一个想法，能否准确地定性一个IP到底做过什么样的坏事儿?

什么是坏事儿？

反欺诈中，涉及到的业务风险其实非常非常多。不同的行业、不同的平台都会有各自独有的一些风险。

就拿"黄牛"来说，随着互联网的发展，黄牛也从最早的票贩子，演变出了很多很多的花样。

案例1:在各大航空公司的网上订票渠道中，存在很多"占座黄牛"，他们通过特定的渠道，订购了一定数量的廉价机票，然后加价转售，甚至会高出这张机票原有的价格。如果

案例2:一些票务网站(专指演唱会、赛事门票)，黄牛会注册大量账号，抢购演唱会门票，拿到门票后，加价出手。由于黄牛拿到了实体票，转手过程是在线下进行的，通过线

案例3:美团、猫眼、格瓦拉等购买电影票的平台中，也存在很大数量的黄牛。尤其是一些热门大片儿的首映票，价格可以炒到很高。电影票的黄牛，往往以代购的形式操作，

如果我们需要分析一个IP到底做了什么坏事，就必须要先给出明确的定义，到底什么样的行为算是坏事。然后把这些行为分解为非常详细的特征指标，进行建模。

这个过程是漫长的，就像上面举的例子，同样是"黄牛"，由于不同的平台，不通过的行业类型，中间存在着非常巨大的差异。每一种行为都要做这样的深入分析和研究，其实

在后来的一段时间里，我们团队接到了越来越多的提问，客户希望知道，这个IP到底干了什么?到底有没有风险?我们只能硬着头皮，去提取这个IP在过去半年里的行为数据，

首先，我们梳理了一份反欺诈的词表，用来给出各种欺诈行为的明确定义。

特征提取

篇幅有限，这里就简单介绍一下我们对黄牛(票务行业)做行为分析和建模的过程。

A	B	C	D	E	F	G	H	I
注册手机号	注册时间	用户名	手机号	省份	城市	区域	详细地址	下单时间
15546[redacted]640	2016-07-23 05:40:47.0	欧阳我的	15546[redacted]640	广东	佛山市	顺德区	均安镇均根路天连大道10115815987083	2017-02-22 11:18:07.993
15088[redacted]655	2016-07-23 05:34:36.0	欧阳里	15088[redacted]655	广东	佛山市	顺德区	均安镇均根路天连大道1011小区2号楼3单元504室13450862863	2017-02-22 11:18:03.017
13464[redacted]515	2016-07-23 05:37:44.0	欧阳孤度	13464[redacted]515	广东	佛山市	顺德区	均安镇均根路天连大道1013450862863	2017-02-22 11:18:03.533
15269[redacted]945	2016-07-23 05:34:05.0	欧阳一	15269[redacted]945	广东	佛山市	顺德区	均安镇均根路天连大道151小区2号楼3单元504室13652535414	2017-02-22 11:18:03.477
13612[redacted]398	2016-07-23 05:35:16.0	欧阳一台	13612[redacted]398	广东	佛山市	顺德区	均安镇均根路天连大道191小区2号楼3单元504室15815987083	2017-02-22 11:18:03.587
13662[redacted]379	2016-07-23 05:52:05.0	欧阳衬款	13662[redacted]379	广东	佛山市	顺德区	均安镇均根路天连大道191小区2号楼3单元504室18033496499	2017-02-22 11:18:03.017
15084[redacted]733	2016-07-23 05:28:48.0	欧阳略	15084[redacted]733	广东	佛山市	顺德区	均安镇均根路天连大道1号1小区2号楼3单元504室18680041128	2017-02-22 11:18:03.077
17087[redacted]343	2016-07-23 05:29:44.0	欧阳昭昭	17087[redacted]343	广东	佛山市	顺德区	均安镇均根路天连大道2013652535414	2017-02-22 11:18:12.847
15653[redacted]664	2016-07-23 05:45:51.0	欧阳书店	15653[redacted]664	广东	佛山市	顺德区	均安镇均根路天连大道2151小区2号楼3单元504室15815987083	2017-02-22 11:18:03.373
15230[redacted]745	2016-07-23 05:46:58.0	欧阳欧阳	15230[redacted]745	广东	佛山市	顺德区	均安镇均根路天连大道2318680041128	2017-02-22 11:18:02.907
13413[redacted]239	2016-07-23 05:44:24.0	欧阳一台	13413[redacted]239	广东	佛山市	顺德区	均安镇均根路天连大道2613809685306	2017-02-22 11:18:03.627
15230[redacted]859	2016-07-23 05:36:49.0	欧阳基地	15230[redacted]859	广东	佛山市	顺德区	均安镇均根路天连大道271小区2号楼3单元504室15815987083	2017-02-22 11:18:03.657
15940[redacted]763	2016-07-23 05:30:31.0	欧阳热	15940[redacted]763	广东	佛山市	顺德区	均安镇均根路天连大道2913450862863	2017-02-22 11:18:03.45
18242[redacted]062	2016-07-23 05:30:57.0	欧阳热	18242[redacted]062	广东	佛山市	顺德区	均安镇均根路天连大道3013809685306	2017-02-22 11:18:02.947
15679[redacted]453	2016-07-23 05:40:15.0	欧阳热	15679[redacted]453	广东	佛山市	顺德区	均安镇均根路天连大道301小区2号楼3单元504室13450862863	2017-02-22 11:18:03.487
15541[redacted]554	2016-07-23 05:33:39.0	欧阳欧阳	15541[redacted]554	广东	佛山市	顺德区	均安镇均根路天连大道731小区2号楼3单元504室13450862863	2017-02-22 11:18:03.397
15504[redacted]271	2016-07-23 05:51:36.0	欧阳荣福	15504[redacted]271	广东	佛山市	顺德区	均安镇均根路天连大道813652535414	2017-02-22 11:18:03.483
15604[redacted]377	2016-07-23 05:44:52.0	欧阳里	15604[redacted]377	广东	佛山市	顺德区	均安镇均根路天连大道8318033496499	2017-02-22 11:18:03.503
18675[redacted]458	2016-07-23 05:43:57.0	欧阳涉进	18675[redacted]458	广东	佛山市	顺德区	均安镇均根路天连大道831小区2号楼3单元504室13652535414	2017-02-22 11:18:03.47
13969[redacted]442	2016-07-23 05:39:19.0	欧阳里	13969[redacted]442	广东	佛山市	顺德区	均安镇均根路天连大道9313809685306	2017-02-22 11:18:03.24

上图中，是我们抽取到的一份较为典型的黄牛抢票记录。

从这些记录里，能获取到怎样的信息呢?

1、这批账号都在同一天注册，并且注册时间较为集中，注册时间间隔大约为30秒;

2、每个账户只下一个订单，但是多个订单产生的时间非常接近，时间间隔仅为毫秒级;

3、多个订单中的收货人姓名很相似，直观判断，不太可能是真实的姓名;

4、多个订单中的收货地址有明显的异常，在末尾添加了无用的字符串;

5、收获地址末尾的字符串为11位的数字，比较像手机号，多个订单中的这个字符串相同;

6、账号注册和风险发生，中间存在较长的时间，可以定义为休眠账号或养号行为。

如果对这个地址做检查，我们会发现：广东省佛山市均安镇均榄路天连大道是真实存在的。

但是这附近并没有什么小区，反而更像是一个村子。也就是说，收货地址中，“天连大道”之后的部分都是随机添加的，可能并没有任何意义。

这样的做法，是为了避免平台对收获地址做校验，如果大量订单都寄送到同一个收获地址，那么这些订单都存在刷单的嫌疑。

上面的地图中，你可能也注意到了，其实并没有“天连大道”和“天连路”，其实是同一条街。但是由于名称不同，在地址核验过程中，就有可能被认为是两个不同的地址。类似

为此，我们也建立了一套用于对收货地址做真实性核验的系统，用于判断多个地址，是否指向了同一个地点。

除了前面列举的三个特征之外，还有一个比较隐蔽的特征，就是注册这些账号的手机号，其实都是”虚假号码“(参见:互联网黑产剖析——虚假号码）。换句话说，提交这些订

以上种种，就成为我们判断黄牛行为的特征，归纳如下:

1、黄牛会事先通过垃圾注册准备一批可用的账号，注册过程中往往会使用虚假号码；

2、账号注册过程中会出时间、IP、设备上的集中性，即同一个设备，同一个IP上注册了大量账号；3、多个订单中的收货人、收货地址不真实或相似度极高；

4、多个订单可能从同一个设备上产生；

5、提交订单的IP地址，大部分是机房IP或者代理IP；

6、垃圾账号注册完成之后可能不会立即进行抢票，而是经过了较长的沉睡期或进行特定的养号活动……

进一步细化之后，得到具体的指标参数，就可以进入训练模型的阶段了。

攻击链路

攻击链路(aka
Kill-Chain)，是安全领域中一个讨论比较多的话题。任何一次风险，都不会平白无故地发生，而是会有一个过程。对一次风险的定义，可以从最终的结果进行定义，但是更多

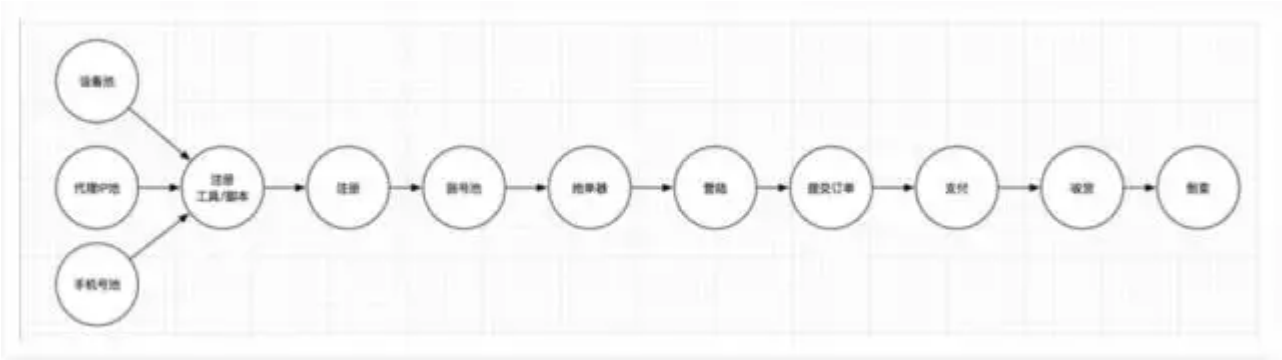
以偷窃为例，一定会有这么几个步骤:

寻找目标 -- 蹲点 -- 标记 -- 作案准备 -- 撬门/扒窗 -- 进入房间 -- 寻找保险箱 -- 撬开保险箱 -- 拿走钱/珠宝 -- 清理现场 -- 离开现场 -- 销赃 -- 寻找下一个目标

上面的这些，就是Kill-Chain中的节点(Node)，也可以叫做风险过程(Process)。在整个攻击链路中，只有起点和终点是确定的，剩下的部分，可能会没有，也可能因为各种实

欺诈风险，也是一样的。前面分析黄牛的特征中，我们提到了黄牛会使用一批垃圾账号进行下单。分析一个账号的欺诈行为，需要纵观这个账号的整个生命周期，或者在既定

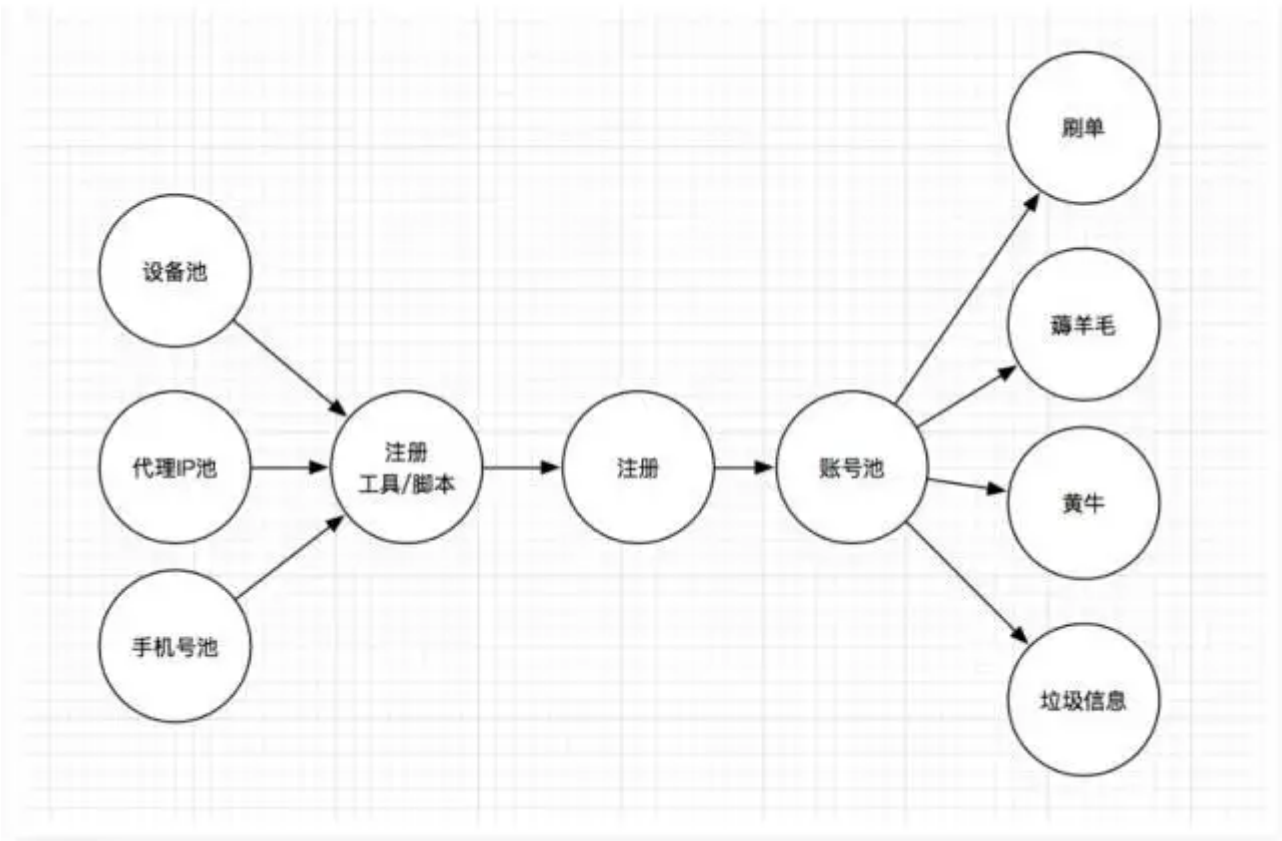
那么，针对黄牛风险，攻击链路就可以表示如下:



在攻击链路中，越是靠前的节点，发现和识别越为困难，因为各种特征其实并不明显，只能判断本次事件有嫌疑，而不能确定具体的风险。但是在这些环节上进行防护，起到

越是靠后的节点，发现和识别变得简单，很多特征都比较明显，但是防护就变得困难。并且，由于攻击链路本身会产生很多分支，可能在其他环节上已经产生了，即便是同一

此外，某些节点上会产生大量的分支链路，比如垃圾注册。通过注册工具/脚本，批量产生的垃圾账号，可能在后续的多种业务场景中出现，不同的业务场景中，又有着不同



平台的业务越丰富，这个分支就会变得越发明显。如果一个平台同时提供了电商、电影票、团购、点评等多种线上业务，那么这个攻击链路就会变得非常复杂。

这也是为什么我们要建立IP地址画像、手机号画像和设备画像的原因。通过已知的各种风险行为，建立模型，通过跨平台、跨行业来进行联防联控，只要这个手机号、IP或者

在整个攻击链路最开始的地方进行防护，并且在账号的整个生命周期中，进行持续监控，使得最终能够造成风险的账户数量降至最低。

在对抗中进步

这场欺诈和反欺诈的对抗，已经持续了多年，并且还将继续下去。

我们在不断提升检测能力、改进检测方式的同时，欺诈分子也在不断地产生新的作弊手段。并且，互联网在不断地寻求创新，同样是促销活动，在不同的平台上，会有截然不

一旦新的业务模式产生，欺诈分子也会相应地寻找可供利用的业务逻辑缺陷，甚至产生一些新的风险类型。这需要我们不断地观察、学习和改进。为此，我们引入了无监督模

结语

我们识别出的每一次风险行为，都会作为标签，标记在手机号、IP和设备上。即使欺诈分子不断地更换这些信息，也总会被发现出来。这是同盾跨行业、跨平台联防联控的目

这些标签，目前在IP画像中已经可以使用，随着我们研究的进一步深入，越来越多的模型被开发出来，可以准确识别的风险行为也越来越多，力求让欺诈分子无所遁形。

点击收藏 | 0 关注 | 0

上一篇：【反欺诈专栏】关于IP，这里有你想... 下一篇：【反欺诈专栏】关于IP，这里有你想...

1. 0 条回复
   - 动动手指，沙发就是你的了！