黑客派

互联网广告作弊十八般武艺 (上)

作者：fjun

原文链接：https://hacpai.com/article/1530701153794

来源网站：黑客派

本文原载于“计算广告”公众号，作者曾宪超、北冥乘海生。经过深入采和精心准备，我们毅然决定做“互联网流量作弊的秘密”这期 live。Live 里的内容绝对是干货，学到些知识，你可以毅然投身“流量优化”行业，也可以在自己业务中擦亮双眼避免上当。

中国在线广告的从业者，都有一颗感恩之心：**不论广告主给了你什么样的 KPI，不论你的流量么不堪，都会创造条件把 KPI 完成，有人把这戏称为“作弊”。**当然，除了有感恩之心，还必须要工匠精神。为了帮助大家在这个行业顺利发展，我们与秒针营销科学院一起，悉心整理了一个合格的告人应该掌握的十八种常用手段，姑且称为“作弊十八般武艺”吧！带上这十八般武艺，在精准营销大数据的道路上坚定前行吧！
在正式介绍十八般武艺之前，我们先来快速过一下广告投放的全过程，看看作弊都可能存在于哪环节。

广告主与媒体或代理商签订广告合同，约定结算方式并提供广告创意。主要结算方式有：按展示结算(Cost Per Mille, CPM)和按点击量结算(Cost Per Click, CPC)的手段做为一类；将按转化量结算(ost Per Action, CPA)和按销售额结算(Cost Per Sale, CPS)。
广告市场中往往有第三方来监测广告效果，保障广告主的投入产出比。如果是 CPM/CPC 结算，第三方会在广告展示环节添加检测代码/SDK，随着广告一齐到达用户端；如果是 CPS/CPA 结算的第三方会在广告主网站或应用内添加检测代码/SDK，进行效果归因(Attribution)。
媒体展示广告，用户看到广告创意。
用户产生广告交互行为(展示、点击、下载和注册等)。在第三方代码的控制下，这些行为连带用信息一齐被发送到第三方，第三方进行统计。
第三方将统计得到的数据报表交给广告主，广告主凭借这份数据与媒体按照指标进行结算。

看起来，广告的逻辑和流程都挺透明公开的，似乎没有什么可以作弊的地方，究竟是哪里出了问呢？近现代史老师告诉我们，凡是遇到问“根本原因”的选择题，只用在 ABCD 里找有“资本”字眼答案即可。在广告中，要想理清业务的脉络，跟着钱的流向走准没错。以 CPM 为例，广告主按照第方提供的曝光数据与媒体进行结算，而第三方的数据来源于用户端接收到的广告展示，广告展示又是过第三方的检测代码统计来的。**从数据到展示，从展示到检测代码。只要检测代码认为广告确实被示了一次，那么不管该用户是否真的见到了广告，广告主都要为此次曝光付费。所谓作弊，就是一个代码说谎的手段。**那么问题来了，如何能让检测代码说谎呢？这就是咱们要聊的“作弊十八般武艺呀！
作弊手段与广告主要求的 KPI 有直接关系。从方法论来看，可以将作弊手法分为两类：**针对 CM/CPC(记为M)**的手段为一类；**针对 CPA/CPS(记为 S)**的手段为另一类。
从另一个角度，还可以将广告作弊分为虚假流量作弊(记为N)和量归因作弊(记为A)。所谓虚假流量，也称为 Non-Human Traffic(NHT)，指的是广告的示、点击或转化本身就是伪造出来的；而后者，则是将其他渠道的流量或者自然流量记在自己名下。般来说，CPA/CPS 的广告由于伪造转化的成本较高，多采用归因作弊的思路。
另外，还可以根据作弊的手段，将广告作弊分为机器作弊(记为 R)和人工作弊(记为 H)。相比之下，机器作弊十足炫酷、易规模化，具有大数据和人工智能科技感觉；而人工作弊则精巧细致、韵味悠长，让人回忆起故乡醇厚的雾霾味道。

<ins class="adsbygoogle" style="display:block" data-ad-client="ca-pub-5357405790190342" data-ad-slot="8316640078" data-ad-format="auto" data-full-width-responsive="true"></in>
<script>
    (adsbygoogle = window.adsbygoogle || []).push({});
</script>
为了方便读者贯彻落实广告作弊的精神，我们将十八般武艺分成上下两篇来进行介绍。今天，咱先来看看 CPM/CPC 作弊手段。
一、直接访问监测代码 (M,N,H)
监测代码是指那些具有客户端信息收集功能的代码。它的主要工作，是将客户端的信息以参数的式拼凑成 URL，并以 HTTP 请求的方式传给第三方，告知“谁，在什么时候，看到了来自哪个媒体展的，哪个广告主的广告”。以移动端为例，常见的客户端参数有如下几种（数据来自《中华人民共和广告行业标准》）：

除了这些，常见的需要被收集的参数还有展示广告时间戳、操作系统、浏览器、设备类型、联网式、APP 信息和标准 UA 信息等。当广告在客户端产生了曝光，监测代码就会记录此次曝光，并采集户信息，其生成的 URL 如下所示。除了第一个参数使用“?”连接外，后面参数都用“&”连接从这段 URL 中很容易读出几个信息：用户的 IP 地址是 10.26.78.45，使用设备 UA 是 iPhone，IDFA 是 70E0E6465B7B12C844C63EC681C7507C 等。直接对这个 URL 发起 HTTP 请求，第三方 http://www.xxxxxcom 就可以根据 URL，解析出广告、媒体和用户的三方信息，在后台形成日志，作为一次正常广告曝光。在行业中，常说的“检测代码”指的就是这个检测 URL，而非装填 URL 的代码，本文亦此。
<blockquote>
 http://www.xxxxx.com/mp?CID=ad20&CPID=1321&CRID=20&OS=1&IDFA=70E0E6465B7B12C84C63EC681C7507C&OpenUDID=F1C7976BC455CB548BFC550EB7687F06&IP=10.2678.45&UA=iPhone;%20CPU%20iPhone%20OS%206_1_2%20like%20Mac%20OS%2X)%20AppleWebKit/536.26%20(KHTML,%20like%20Gecko&TS=1198628984102
</blockquote>
既然是个 URL，严谨的大数据从业者一定会思考：直接在浏览器地址里输入这段代码，是不是也在广告主那里记录了一个曝光呢？是的，这就是作弊刷量最朴素的哲学原理。这其实算不上什么武器只是个玩具，我们借此来说明基础的 CPM 作弊原理，CPC 也是一样一样的啊！
二、服务器刷监测代码 (M,N,R)
靠手工输入监测代码的方式来刷曝光虽然原理上可行，实际生产中则是没有什么卵用的，量太小不足以产生质的影响。那能不能写一个爬虫程序，自动装填各种参数，自动发起 HTTP 请求呢？咳咳你看看，这人要是想学坏，是真容易呀！可以租一些云服务器，把代码都搬到天上去，一键云作弊。
对于服务器刷代码的作弊手段，第三方是蓝瘦香菇的，占用了服务器大量带宽不说，虚假流量的入为真实效果的统计也提出了严峻的挑战。当然，服务器刷代码的方法还是有漏洞的：云机房的 IP 址大多属于同一 IP 段，屏蔽掉主要云服务提供商的 IP 段即可——谁没事吃饱了撑的租了服务器上去广告呢？
<script async src="https://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js"></script>
<!-- 黑客派PC帖子内嵌-展示 -->
<ins class="adsbygoogle" style="display:block" data-ad-client="ca-pub-5357405790190342" data-ad-slot="8316640078" data-ad-format="auto" data-full-width-responsive="true"></in>

三、客户端刷监测代码 (M,N,R)
用服务器刷监测代码， 虽然简单直接，却在 IP 和 cookie 等用户身份统计上很难做到自然。于，勤劳勇敢的作弊人们又想到了一个新办法，直接在客户端刷监测代码。您觉得访问了一个网页，其网页上的 JS 又免费赠送了您好几次浏览，或许还有一次点击。这样一来，从用户行为上就很难找出么马脚了。
当然， 这样的作弊也不难发现：上次我偶尔看到某汽车网站一次广告投放的用户频次，大多数在 8/16/24/32 这些吉利的数字上。这是为什么呢？就是给用户的正常浏览都买一赠七了呗！如何自化找出这样的作弊呢？我说两个关键词：傅立叶变换、频域，懂的码农自然懂了，不懂得恐怕还要去习一下《信号与系统》，这超出了本文的范畴。
另外，不论是服务器刷还是客户端刷，在点击环节都会有个破绽：正常用户在点击广告时，自然点击分布与广告创意有关，而刷的点击要么较为集中，要么均匀散布，并不难以分辨。画个点击热力，就一目了然了。

四、频繁换用户身份 (M,N,R)
广告投放中的用户身份，不会是 email、手机号等 PII 信息，一般情况下，在 Web 场景下用 cookie，在苹果手机原生应用中用 IDFA，在安卓手机原生应用中用 AndroidID，如果这些都没有，就用 FigerPrint(IP + User Agent)。
不论您采用哪种刷量的手段，一般来说都要比较频繁地变更用户身份。否则，在一个 cookie 身猛薅羊毛，一个用户有成千上万次展示，一看就知道是假的。因此，频繁换用户身份，是作弊行业的本功之一，也是反作弊时都应该了解的一点。
这个方式怎么对付呢？对可以选流量的 DSP 来说，有个简单的办法：凡是第一次看到的 cookie 或设备，就干脆不要出价了。
不过对供给方产品来说，这个法子就行不通了。但是，供给方有供给方的好处，对于移动上以 SK 方式潜入媒体的 SSP 产品来说，可以拿到很多终端的信息。举个小例子，如果一台手机的电量总是的，十有八九是有问题的，您明白了么？
五、放 iframe 造假展示 (M,A.R)
iframe 是一个 HTML 标签，可以在当前页面中插入其他页面的内容，常常被用来作为承载展示告的载体。iframe 有诸多属性可以设置，其中最为广告人喜爱的莫过于宽度 width 和高度 height 了通过对这两个参数的设置，可以将广告尺寸从肉眼可见的 220140 变成不可见的 11。就改了一个大小嘛，这算什么作弊呢？各位，大小改了之后，虽然你看不见，但检测代码看见了呀，是一次正常的广告展示，可以向广告主收钱了。可见，作弊者对我们广大用户还是有一颗怜悯之心的我就挣广告主的钱，向借您一个像素点，广告不会显出来，不影响您正常上网。这不正印证了那句老嘛： “不打扰，是我的温柔” 。
在展示广告中使用 iframe 的手段进行作弊，广告效果自然是很差的，用户并没有看到宣传，广主白花花的银子就没了。除了展示广告之外，视频广告也面临着类似的问题。例如，在一些新闻页面会在一个非常不起眼的位置上播放视频广告，效果也是极差的。

六、肉鸡和 Root(M/S,N/A,R)
“肉鸡”是指那些被木马感染，可以被黑客远程控制的机器和设备。一说到黑客，大家可能就觉“哇噻好厉害”，其实一点都不难。不知道大家有没有听说过“灰鸽子”，某人在小学 4 年级时是一

script kids，而且还成功的捉到过一只肉鸡。就在要远程登录的一瞬间，肉鸡不见了，应该是被杀毒件 Kill 掉了，从此卸甲归田从了良。言归正传，这个鸡呀，噢不，肉鸡，它就是一个正常的用户，可发起浏览和点击等行为，因为肉鸡的后面是真人。提醒各位，注意上网安全，小心被捉鸡，指不定会什么乱七八糟的东西呢。
Root 是指操作系统中超级管理员权限，当拿到 Root 权限后，整个系统就是你家，拆了都可以这里所说的 Root 访问广告，主要指的是在移动端，某些 APP 获得了 Root 权限，就可以在后台静悄的进行着各种各样的访问、点击和下载操作，也都是真实的数据。与肉鸡不同的是，肉鸡后面是有真在操作，而 Root 更多的是程序在执行，但从效果上来看，都是在用户不知情的情况下，在后台进行各种各样的广告操作，欺骗第三方检测代码产生真实的用户行为数据。
说到这个 Root，可以说是移动时代“效果广告”的神器！去年，有一家中国公司的 Root 程序至惊动了美国 FBI，差点被当成窃取美国用户信息的典型而破坏中美友好大业，其实人家只是很单纯想挣点儿广告费；而另外一家上市公司（请不要问我是哪一家）则收购了业内一家 Root 变现大师级业，被他们 Root 的手机，除了 24 小时弹广告之外基本上就没啥用了，据说当天就能回本儿，但是户的留存就很差了，为啥？连手机都摔了！
七、诱骗用户点击广告 (M,A,R)
诱骗用户点击广告的情况通常发生在 BBS 中，经常会有广告伪装成帖子，以博人眼球的内容诱用户产生点击。在点击之后，就会发生页面跳转或者触发软件的下载，用户很容易中招。除了 BBS 外，在有些下载网站中，有很多的“点击下载”，相信各位也都经见过，点击之后指不定就跳到哪里了，反正我是被带到某特卖网站上转了一圈，看到 0.2 折的貂绒大衣正准备剁手，才想起来原来我是下载软件的。

从上述这些 CPM/CPC 的作弊手段中可以看出，广告作弊的一般思路都是围绕着检测则而进行的各种 Hack，用虚假或低值的流量完成订单，骗取广告主的预算。只要抓住这本质，广告的作弊手法就不难理清了。至于作弊者的底线是什么，不好意思，恐怕连底裤都不知道是物。
《互联网广告作弊十八般武艺(上)》到这里就结束了。在下篇中，我们将就 CPS/CPA 作弊手段行介绍，花样之繁多，手段之丰富，恐怕会令您咋舌。不信？在下篇里，我们将介绍运营劫持、Cookie Stuffing、游戏自冲、淘宝代销、Cloaking、下载归因等五花八门的作弊段，我们下篇不见不散，科科