设备指纹

1、设备指纹(Device Fingerprinting)是什么？

简单来讲，设备指纹是指可以用于唯一标识出该设备的设备特征或者独特的设备标识。

设备指纹包括一些固有的、较难篡改的、唯一的设备标识。比如设备的硬件ID，像手机在生产过程中都会被赋予一个唯一的IMEI(International Mobile Equipment Identity)编号，用于唯一标识该台设备。像电脑的网卡，在生产过程中会被赋予唯一的MAC地址。这些设备唯一的标识符我们可以将其视为设备指纹。

同时，设备的特征集合可以用来当做设备指纹。我们将设备的名称、型号、形状、颜色、功能等各个特征结合起来用于作为设备的标识。这就类似于我们在记忆人的时候，通常是通过人的长相、面部特征来记忆。

**设备指纹背景介绍**

在用户访问网站的过程中，网站需要跟踪用户在当前页面执行了哪些操作。比如搜索了哪些关键词、浏览了哪些商品，该设备的用户今天尝试登录了多少次，注册了多少个账号等等。

1.基于cookie的用户追踪

HTTP是无状态协议，cookie技术的产生使得用户可以将相关的状态存储在本地客户端，进行会话保持，同时使得服务端能够知晓用户当前所处的状态。通常是访问网站时，网站生成一随机的唯一标识id返回给客户端存储。客户端在之后的每次请求中都会携带该标识符。用户在当前页面执行查询、购买、评价等行为时候，服务端都会将行为进行记录。

2.持久化cookie

针对很多cookie会被清理的问题，为了使得cookie更持久，用于用户追踪，就产生了像evercookie类似的技术，其不仅设置cookie，同时会将cookie的值存储在Flash Local Shared Objects或者HTML5 Storage、IndexDB、IE的userData Storage等位置，其甚至还会利用JAVA的安全问题，比如CVE-2013-0422绕过applet沙盒，然后将cookie存储到用户的本地文件里。evercookie的具体实现过程可以访问其[项目文档](https://github.com/samyk/evercookie)，此种技术产生已久，现在的垃圾清理基本可以做到完整清理cookie了。

3.无cookie的用户追踪

由于cookie存储于客户端，可被清理，同时目前浏览器都会提供隐身访问模式，在关闭浏览器页面后cookie等信息会被自动删除。基于cookie的用户跟踪还会有隐私问题，无cookie的技术应运而生。

Mayer(具体可参照其paper，Any person... a pamphleteer” Internet Anonymity in the Age of Web 2.0)在2009年以及Eckersley(具体可参照其paper，How Unique Is Your Web Browser?)在2010年都曾提出基于浏览器的特征以及插件信息，不依赖cookies进行设备的识别。

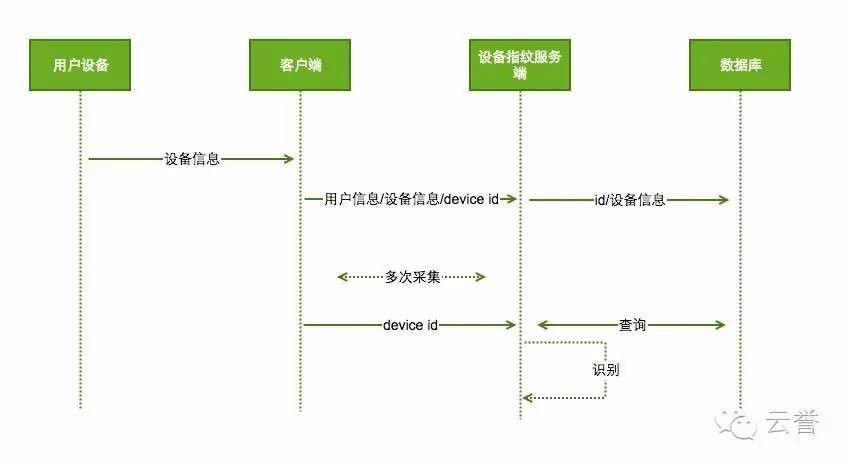
由此基于无cookie的设备识别技术就产生了，即我们通常所说的通过设备指纹来进行设备识别、用户追踪。设备指纹识别系统则是通过收集众多的设备指纹信息，通过指纹信息定位设备，从而跟踪其在网站的行为。

**设备指纹的实现原理**

日常生活中我们通过姓名来区分人，在辨别或记忆人时通常是通过他的相貌，面部特征来进行。设备指纹的实现原理也类似，其通过获取设备的各项信息，然后综合各个特征，按照一定算法形成该设备的标识符(即设备指纹)用于区分识别设备。

**实现流程**

流程图：



说明：

设备指纹的实现过程，通常是通过用户的客户端，比如web浏览器、app应用、pc软件等，将用户的设备信息提取，按照一定的hash算法计算获得一个唯一id，同时如果某服务包含业务数据，那设备信息通常是与业务数据进行关联的，并一同发送至设备指纹服务端进行处理并存储。设备指纹服务端在收到用户的数据，按照自己的存储方法，算法进行存储关联。当然由于用户的设备信息可能会有变更，比如ip地址等，所以会伴随多次的信息采集，多次的设备id计算。

在随后的业务请求中，通过设备指纹识别系统的算法进行设备的识别与关联。

**信息采集**

设备指纹识别系统的实现必然伴随信息的采集。主要分为主动和被动的信息采集。

被动的信息采集在发起网络请求，一些基本的属性会自然携带。比如用户的IP信息、端口、请求头以及User-Agent等信息。

主动的信息采集

是指在设备上执行特定的代码，来采集所需的设备信息。比如通过代码采集设备的mac地址、操作系统版本、CPU型号等信息。

其中基于浏览器的设备指纹采集较为常见。浏览器的设备指纹通常需要如下一些内容，通过在浏览器执行js或在flash代码，获取设备信息。比如UserAgent、时区、屏幕大小、浏览器插件、系统安装字体、浏览器的canvas的特征、浏览器的webgl特征等等，具体实现我们将在之后的文章中阐述。

**设备指纹的应用场景**

1.行为追踪

用户行为追踪主要与业务相关。比如购物类的网站其会采集用户的设备信息，并根据设备指纹信息对用户进行相关的商品推荐。当然类似商品推荐只是一方面，更多的是对于采集的信息形成设备指纹，给用户提供较好的业务服务以及安全保障。比如检测到用户的风险登录、更换设备登录会要求用户进行较强的二次验证。

2.广告推广

是指结合用户的搜索记录、浏览记录等，对设备进行记录，针对性地推送广告。对于清理cookie、开启广告block、禁止追踪的情况，基于设备指纹识别的技术可以使得广告推送同样精准。

3.反欺诈

其实设备指纹在反欺诈风控中扮演着重要角色。通过设备指纹系统可以给相关业务提供安全保障。比如现在风险较高的垃圾注册、盗号、撞库、异地登陆等异常的行为，采用设备指纹识别系统可以得到有效的控制。

    1）点击欺诈

     点击欺诈，主要是应用在广告作弊方面。广告方按照用户点击广告的次数提供佣金。基于指纹的识别技术可以统计出相同设备点击广告的次数，找出广告作弊的请求，从而可以减少点击欺诈。

    2）金融反欺诈

     我们在[金融反欺诈-海外信用卡黑色产业链](drops.wooyun.org/news/14382)一文中简单提过设备指纹，其在信用卡反欺诈中扮演着重要角色，是风控模型中的一个重要维度。同时在金融交易中，需要确保当前交易在一可信的设备上，这依赖设备指纹识别系统。

    3）防薅羊毛

     薅羊毛的场景一般如下，羊毛党批量注册用户，并领取实物、优惠券、返现券等。批量注册行为一般伴随代理IP的切换，但设备可能不会变化。所以可以基于设备指纹进行设备识别，对同一设备注册账号、参与活动次数进行计数，并设置相关的防控规则。

**总结**

设备指纹有着广阔的应用前景。伴随互联网诞生有一句经典的论断“在互联网上，没人知道你是一条狗。”然而随着互联网的发展逐渐成为人类的第二空间，网民无法识别身份、无法自证信誉已经极大的阻碍互联网业务的拓展。

现实社会有着成熟的征信系统，对于一般安全需求的场景可能仅需要用户用身份证来自证身份即可准入，对于高风险场景（如贷款）甚至需要征信记录才能准入。那么互联网场景是否可以建设成熟的征信体系呢？

互联网征信体系建设有两种思路，一是直接将线下的征信迁移至线上，一是既然互联网已经是一个成熟的生态，那么直接基于用户在互联网的活动记录进行信誉评分。而设备指纹即是互联网信誉评分的基石。

目前，基于设备指纹的的风险防控服务已经逐渐成熟。 在国外,比较成熟的商业化公司有ThreaMetrix,该公司有着海量的设备信息,并拥有自己的设备识别算法与设备信誉评级。国内的像阿里巴巴云誉反欺诈,基于阿里的大数据,通过自有的行为收集和机器学习模型,提供成熟的设备信誉、风控服务。

**2、设备指纹现状**

主动式主要通过SDK或JS代码在客户端主动收集设备信息，来实现设备的精准识别，响应速度和准确度相对较高，但使用场景常因为隐私保护而受限。

    与之相对，被动式主要通过在服务器端收集通信协议和网络的特征来识别设备， 100%保护用户隐私，因而有更大的适用范围，尤其是一些对隐私信息非常敏感，不便于植入SDK或JS代码的场景，如互联网金融行业。同时，被动式设备指纹也有着算法和系统复杂度高、响应时间较长、研发难度大等局限。

      国际上早已有“no cookie”技术相关的政策法规，以确保设备在线追踪的合法性，例如在欧盟cookie原则下，企业在追踪消费者行为前，被要求向消费者请求许可。在美国，广告自律委员会(ASRC)的问责项目近日也向其强制项目的多项需求中添加了“无cookie技术”。针对此类相对更为复杂的追踪技术，企业必须向消费者提供一个拒绝接受的选项，以保持合法性，避免强制执行。在国内，相关的数据隐私政策法规也与国际接轨，愈加完善。

     未来，主动式设备指纹技术的应用场景会受到越来越多的限制，而被动式设备指纹技术则是大势所趋。

    上一期最后我们也提到，除了主动式和被动式，实际应用中还有第三种设备指纹技术，那就是混合式设备指纹识别技术。

    混合式设备指纹技术指将主动式和被动式设备指纹技术整合在同一个设备识别与跟踪的架构中，将主动式设备指纹技术在客户端生成的设备标识符，与被动式设备指纹技术在服务器端收集的、协议栈相关的特征信息对应起来，使得所有的设备都有一个唯一的设备识别ID。

     混合式设备指纹技术融合了主动式和被动式设备指纹技术各自的优点，在准确识别设备的同时，扩大了设备指纹技术的适用范围。

准确率进一步提高：

    主动式设备指纹技术本身已经具有很高的准确率，接近100%。在此基础上，混合式设备指纹技术能够将主动式和被动式的识别结果相互印证和校验，从而使准确率进一步提高。

适用范围更广：

    出于隐私保护的目的，纯主动式设备指纹技术的使用场景越来越受限，同时其对于SDK和JS代码的依赖，致使其无法应用在不便于在客户端植入代码的场景。例如，Apple公司就严格限制APP中用户与设备信息的收集行为，任何超越合理范围的信息收集都可能被视为违规行为，而遭到APP下架的惩罚。

     而混合式设备指纹技术则可根据用户的不同场景，来灵活决定设备指纹的生成和校验方式。对于用户隐私保护要求宽松的场景，可以使用主动式设备指纹技术来获得快速、稳定的设备标识；而对于用户隐私保护要求严格，或无法嵌入SDK和Javascript代码的场景，则可以使用被动式设备指纹技术。

     此外，在需要将同一用户在移动Web和App中的行为关联起来的场景中，混合式设备指纹技术也有着比主动式更大的应用范围。一个典型的案例是，当一个消费者点击到一个移动网络广告以获取一个新应用时，广告主希望知道此次广告点击是否最终带来了新用户。而由于嵌入Web页面的Javascript代码和移动APP中的SDK收集的设备特征不同，导致生成的设备指纹标识符也不相同。因此主动式设备指纹技术无法将同一用户在移动Web和App中的行为关联起来，而被动式和混合式可以根据相同的协议栈特征，来判断移动WEB和APP发送的消息是否来自于同一台移动设备。

**三种设备指纹技术的现状和未来**

     目前，国内大部分企业采用的设备指纹技术都是主动式的，但由于企业和公众隐私保护意识日渐强烈，隐私保护政策日益严格，主动式设备指纹技术的应用越来越受限。

而被动式设备指纹技术因为系统和算法比较复杂，往往需结合大数据处理和机器学习技术才能实现，技术门槛相对较高，目前在国内还处于起步阶段。只有少数公司可以做到较高的准确率，得到业界和客户的认可。

      混合式设备指纹技术的公司更是凤毛麟角，主要原因之一是由于混合式的技术壁垒也并不比被动式低。能够成功研发混合式设备指纹的公司，往往在被动式上已经有了相当大的进展和积累；或者从外部采购其他公司的被动式设备指纹技术作为混合式的研发基础。在业内较为知名的被动式指纹技术专家如maxent，也已为多家业务伙伴提供此类的技术合作与支持。

    在未来，我们有理由相信，被动式设备指纹技术将逐步成为设备识别技术的中流砥柱。

     而混合式设备指纹技术，作为二者结合、扬长避短的综合性解决方案，同时兼顾了应用场景、用户隐私保护、响应速度与准确性等多方面的要求，可以在智能化的商业新环境中，助力企业在营销、运营、交易等各个环节，实现业务目标。

金融欺诈是以什么形式实现的？好比有人做信用卡套现，自己办了几十张信用卡，利用B卡还A卡后，再利用C卡还B卡的钱，形成了一套微型的庞氏骗局。再比如发起贷款的时候就有一定的恶意的用户，利用程序或者其他手段批量地申请贷款。其实这些欺诈手段与网站优惠注册中羊毛党注册小号批量赚取优惠相仿，并且这些这期手段也能实现“自动化”，以机器人等多种方式自动批量注册用户。

Maxent的创始人张克对我说，在互联网金融的领域，反欺诈主要作用为保护资金账户安全，在贷前、贷中及贷后各环节提供信用欺诈风险识别，也就是尽最大可能排除在与交易有关的环节产生的风险。

Maxent是为互联网提供风险控制和反欺诈实时数据分析服务的创业公司，拥有十几项专利核心技术，而其中的一项底层技术就是被动式全站设备指纹技术，创始人张克对我说，“**通过设备指纹技术等底层技术构建的反欺诈解决方案就能在一定程度上实现：当交易还没有发生时就规避掉这类恶意用户**”。

设备指纹本身也就是指各类智能设备在互联网上的标示ID及其应用轨迹。移动互联网中用于标示用户或设备的ID纷繁复杂，同时由于移动设备更强的私人属性，Maxent要解决的问题就是：如何在不侵犯用户隐私的前提下，能够有效识别用户设备并关联用户行为。因此，**被动式设备指纹技术可以在设备发起贷款的时候，就判断是否恶意设备。**细分的判别标准很多，也要结合本身设备的属性和人的行为的属性做出恶意的风险指数，最后给到B端客户自行进行判断。

通常以行业角度来看，反欺诈解决方案提供商在细分领域互联网金融的目标客户之一为小贷公司。小贷的风险因素主要为集团骗贷，几个人一个团伙，用不同的设备申请小额贷款，虽然每单贷款数额较小，但累计的话可能贷款就是一个庞大的数字。

Maxent通过设备指纹技术等专利技术，**能够识别出异常的贷款设备**，比如单设备单日内贷款申请较多；或者多设备的关联识别为同一用户行为，利用不同设备多次进行贷款，那么Maxent就会将排查出的异常终端设备的信息提交给企业。

如上文所述欺诈等行为的解决方案主要是由Maxent独立开发的Saas平台系统实现的：

猛犸反欺诈Saas服务平台：基于Maxent自身的专利机器学习技术及DFP技术自动发现新的欺诈模式，持续地自行进行反欺诈算法的优化，根据客户交易平台的实时活动事件及历史数据、挖掘海量数据以分辨真实或可疑的设备。

动态实时更新规则：实时监测欺诈行为，主动优化欺诈算法

欺诈指数MoMA Score：提供各类交易的分析页面，将数据简单化、可视化。综合设备IP、地理位置、交易速度等用户数据打分，排查出有风险的设备

自动核算+人工审核：系统自动计算欺诈指数，并且对单笔交易提供接受or拒绝的选项，简化全人工审核到了流程

张克表示，“猛犸反欺诈Saas服务平台具体实用效果为60天内降低人力70%的成本”。而在以上解决方案中Maxent使用多种底层技术共同支撑猛犸反欺诈平台的动态欺诈模式挖掘，建立了各种风险模型，其底层技术的优势主要归结为以下两方面：

设备身份识别：被动式全站设备指纹技术是将同一设备在不同应用、不同场景、不同网络中的行为做关联，产生精准的设备画像，帮助线上交易空风系统定向识别在线欺诈行为。

机器学习：动态地发现欺诈特征，以非监督或者半监督的方式动态调整参数训练的算法模型，由系统自动推断出欺诈风险。在此过程中，当机器需要大量数据不断修正参数去训练模式及算法的时候，根据用户的指令、机器的自我学习，久而久之机器就可以自动计算出用户需要的结果。

也就是说随着单个B端企业用户使用次数的增加，算法进行自我学习的过程中也会进行自我优化，最终随着不同行业给出的数据和反馈，渐渐机器就适应了每个单个的客户，机器计算出的结果会越来越契合用户的业务需求。

机器学习就是让机器学会人识别事物的方法，本身属于人工智能的范畴，包括模式识别、机器学习、搜索、规划等。**机器学习的方法主要有三种：监督学习、半监督学习和无监督学习**。监督学习是利用一组已知类别的样本调整分类器的参数，使其达到所要求性能的过程，简单概括就是根据已知的，推断未知的。半监督方法主要考虑如何利用少量的标注样本和大量的未标注样本进行训练和分类，也就是根据少量已知的和大量未知的内容进行分类。因此，创始人张克及周辉对我表示，**大数据很多情况下陷入了数据的误区，而公司想要解决的问题就是如何更好地利用现有数据为决策做支撑。**

**Maxent主要是应用人工智能半监督式机器学习，利用大量未标注样本进行训练与分类，充分自我训练及自我优化。基于半监督式在线机器学习的全智能评估系统，其计算结果契合客户的业务需求。**同时实时动态地进行欺诈模式挖掘，时刻更新海量信息库，节省人力成本。

**除了机器在人工智能方面的算法自我学习之外，另外一个不断完善、提高数据挖掘、分析准确度的因素就是终端设备的覆盖度，覆盖率越广就越容易做到人机的匹配**。对此，张克认为技术导向的公司很容易就能通过将客户的业绩提升，展示技术能力。

张克的思路是**以产品为驱动**，直观地展示给客户。Maxent提供的Saas服务，力图在免费期内使用户体验到效果。客户提供历史交易的数据，最后得到分析的结果。而**Maxent的技术门槛就体现在：通过分析协议栈里的信息或者通过算法就可以分析出来单用户的多设备关联链接，多种底层技术各有侧重和优势。在不侵犯用户隐私的前提下**仅依靠收集授权及公开的设备信息做数据挖掘、无侵入式追踪，遵守国际隐私法规。

Maxent核心团队移动互联网和广告技术方面的名企高级管理者、技术专家、以及机器学习领域学术界专家构成。团队创建于2014年11月，创始人张克最早在亿动广告传媒的DSP事业部担任总经理，也曾经在思科、朗讯、IBM任职多年，本身从事技术的经历也使他能通过技术的变革发掘行业机会。联合创始人周辉任CTO，曾任亿动传媒大数据首席科学家，1997年起供职于新加坡系统科学院从事互联网网络技术的研究，兼具多年机器学习研究经验，发明高精准的全栈被动式设备指纹技术。联合创始人林晓东为美国Rutgers大学终身教授，是金融工程及记录关联领域的资深学者，另外一位是Prof. Wong，在生物信息查询系统Kieisli及复杂事件处理（CEP）学术领域较有建树，为ACM fellow。