|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | |  | | Macintosh HD:Users:admin:Desktop:finger:finger_logo.jpg | |  | |
| Android Fingerprint  Aim to identify any android terminal |
| chaoranf@jumei.com  2016/11/9 |

# Android Fingerprint

## Aim to identify any android terminal

1. 技术背景及意义

Android在当前智能设备系统中占有的市场份额最多的智能系统，根据Strategy Analytics机构2016年第三季度的统计报告，截止到2016年9月底为止，android系统的全球市场份额已经高居87.5%，远远高于iOS所拥有的12.1%的全球市场份额[1]。

作为一名Android研发人员，看到自己所从事的行业系统发展如此迅猛，除了高兴之余，还出现了隐隐的担心。伴随着行业市场的发展，行业技术也一定会出现很多的变化来适应市场的发展。与我们研发人员息息相关的应该就是Google连续发布的SDK版本号。从2015年开始，Google连续迅速的推出了Android5.0、Android6.0、Android7.0的SDK版本，从系统层面对Android系统进行了大量优化，提供了更多更灵活的API接口，解决或者限制了以前版本存在的安全隐患和性能问题。

在如此市场背景和技术背景更迭的大形势下，有一门技术始终处在技术研究和科学研究的热潮中，它就是指纹识别技术，在我们Android研发领域也被成为Android指纹。它存在目的就是为了唯一识别并标识一台android设备，保证其识别信息与任一其他设备不完全相同。这项技术被广泛使用在广告精确投放、安全认证、访问控制等领域，是一项不可或缺的技术。

1. 技术现状

虽然设备指纹的作用很大，但是目前为止市场上还没有一个统一的技术方案，每个技术团体都有自己的算法和技术实现。一个比较突出的例子就是，在utdid这个属性的使用上，阿里、腾讯、听云、talkingdata、头条、京东等，都使用的自己的utdid算法，来作为自己的设备指纹方案，从表现上来看，主要是为了做设备识别，在安全认证上应该也会有其应用的地方。

聚美utdid的现状，是直接使用imei来做为deviceid的，这是一个很明显的单点瓶颈问题，所以也继续一个技术方案来解决聚美指纹的需求问题。

1. 技术要点

1，设备基础信息的提取

设备基础信息是指一台独立的android手机设备所具有的一些特有参数，本节按照提取方式的不同分开来说明。

第一种是Java层的提取，通常通过android.build等来提取对应的属性信息，该类型的提取特点是易提取、易篡改；

第二种是C层的提取，通常是通过JNI形式来提取对应的属性信息，该类型的提取特点是难提取、较难篡改。

2，设备基础信息的选择

在数据选择的算法中，常见的有FCBF特征选择算法，主要作用就是降维，常用降维方式有小波变换、傅里叶变换、拉普拉斯变换等。

上述算法复杂度对于代码而言都是比较高的，我们引入一种比较简单的概念——信息熵[2]，对于可变参数来说信息熵越小越好，固定不变参数信息熵越大越好。通俗的来讲信息熵表示的是一个数据所代表的信息量。

本文在信息选择上，将信息先划分为两类：不变参数、可变参数，每一类还可以再分为不同的维度，具体细节demo代码中会有对应注释。

首先，对于不变参数，主要选取选择信息熵较大并且较为重要的参数，比如imei可以唯一标示一个手机，但是现在很多手机因为各种原因无法读到该参数，甚至该参数会出现都相同的情况（比如小米某些机型）。

其次，对于可变参数，主要选取在短时间内较大概率会发生变化的参数，比如可以用闪存、可用内存、电池电量等。

3，指纹匹配算法的选取

当取到各基础参数之后，不同的厂商或者商家都会根据自己的需求去设计自己的算法，去匹配自己的指纹方案，以期做到唯一性。常用的指纹匹配算法有关联匹配法，可以在机器学习过程中不断校容错率和准确率的最佳匹配值。

我们初步实现可以使用简单的自定义算法，本文定义的算法为两部分：动态指纹和静态指纹。静态指纹由一系列静态参数以字符串形式连接，然后取md5，动态同样原理只是取的动态参数。在做匹配时，静态参数md5可以作为udid，而动态参数md5主要用来验证合法性。详细逻辑可以看demo或者看流程图。

4，指纹业务算法

业务算法可以包括黑名单、白名单、单个用户持有设备数、单个设备可以登录用户数等，请PM自行设计。

5，整体方案的优化

在使用本文提出的较简单的算法基础上，可以优化的地方主要是指纹业务算法部分。

在使用更科学严谨的特征匹配、关联算法等算法时，则可以通过Matlab等模拟机器学习过程，来不断校正算法规则。

1. demo实现

1，客户端demo

主要实现了信息的采集、指纹算法实现（md5）、提供加盐方式、指纹cookie形式上传；

2，服务端demo

使用Django框架搭建服务器，实现了用户设备信息对应存储、设备信息存储、用户合法性认证。

1. 技术展望

demo已经基本满足了当前对于指纹的需求，可以大大提高技术人员对不同客户端的识别和认证效率，但是如果采用性能更好的算法，则会进一步提高指纹的识别度。

1. 结束语

通过两个多月空闲时间的研究，带来的技术方案可能会存在很多不足，如有问题还请直接提出，我们共同解决。相信我们可以做出更好的方案。

参考文献

1. <http://www.199it.com/archives/532606.html>，中文互联网数据资讯中心

[2] <http://baike.baidu.com/link?url=n3bPRaYPq-y97lysGgir_IsE5-Fp-kRZ_fOqs5_6sdjUK7-cYIVl45zVjrTyjlgpL81xBs1Y8IWkQSe1s25jVgOqz_PB6dPQ0QW0W6pZYvH01ReGoJXRffgmv_b4yBrX,信息熵概念>

[3]Android设备指纹识别技术的研究与实现，研究生论文

[4]<http://blog.csdn.net/zx799959906/article/details/39369543>，特征选择

[5]<http://blog.csdn.net/opennaive/article/details/7047823>，关联匹配