验证码笔记

gitee.com/LongbowEnterprise/SliderCaptcha
https://www.npmjs.com/package/vue-captcha
https://hacpai.com/article/1577444310157
https://github.com/mewebstudio/captcha
https://github.com/EtherDream/proof-of-work-hashcash

一、常见验证码分类

1 文本验证码

以问答形式出现，如：给出问题要求用户答案，计算数字等。

1. 优点：支持广泛，仅需支持文本传输即可应用
2.
3. 缺点：安全性差；所有的验证码问题和答案都要事先在数据库中存好，所以这类验证码数量有限。攻击者可以先将问题库中的所有问题先爬取下来并准备相应的答案库，破解时只需利用正则表达式将验证问题提取出来，然后从答案库中找到相应答案即可。

2 静态图验证码

目前应用最广的一类验证码，这类验证码要求用户输入验证码图片上所显示的文字或数字.

通过扭曲、变形、干扰等方法避免被光学字符识别（OCR, Optical Character Recognition）之类的电脑程序自动辨识出图片上的文字和数字。

1. 优点：实现简单，有比较成熟的方法。可以简单地有效提高攻击者成本。
2.
3. 缺点：随着计算机视觉、机器学习技术的发展。文字识别算法的准确率也越来越高，而过于复杂的干扰线、扭曲也会降低用户体验。

2.5 动态图验证码

对静态图验证码的优化，使用各种GIF甚至flv格式的动画验证码。

1. 优点：未找到相关文献
2.
3. 缺点：事实上提供给攻击者的信息反而多于普通的静态图验证码，识别难度降低。

3 滑块验证码

这类验证码要求用户拖动滑块完成拼图，记录并识别用户的鼠标轨迹从而分辨人机。

前端记录其响应时间，拖拽速度，时间，位置，轨迹，重试次数等。这些因素能够构成一个采样结果或者辨识特性。

通过ajax将其加密并发送至后台判断。

1. 优点：极大优化用户体验。
2.
3. 缺点：识别率存在争议，目前的识别主要是基于对行为信息的简单校验，容易被模拟鼠标轨迹破解。（https://github.com/neuroradiology/InsideReCaptcha）

补充：
该验证码的核心在于加密算法及行为辨识算法，属于行为分析思想的一类验证码。

4 用户行为指纹分析

滑动验证码并不是验证码，它只是一种网页数据加密的方式，其原理是：采集用户的操作数据，环境数据等等，非常多的数据，通过一个算法加密得到字符串，然后提交到服务器分析，服务器有一个判定标准，对数据进行简单的分析就知道是不是人工在操作 — google.recaptcha

淘宝的UA也是从网页加载好了以后就一直记录用户的操作数据，包括鼠标在哪点击，在哪停留，在哪拖拽，键盘输入，浏览器环境，电脑环境等等，大概采集了上千个数据，然后通过一个很复杂的算法加密成一个字符串。

在此方法下，网页首先会收集用户浏览器相关的信息并交付风险评估系统，对不同信用等级的用户提出不用难度的挑战（旨在用户友好）。而这些挑战的主要目的在于增加用户在网页内的操作量，以提供更多的数据进行行为识别。

收集的信息包括：跳转来源；网站的sitekey；浏览器历史和cookie;用户在此页面内的行为。

这些挑战包括：无验证码；复选框验证码；图片识别验证码;文字识别验证码等。

1. 优点：用户体验好，识别精度相对较高，想攻破需要进行复杂的分析。
2.
3. 缺点：所需工作成本大，需要投入成倍的工作在验证算法的更新、升级上。有过使用伪造cookie欺骗系统以跳过验证的例子(Im-Not-a-Human-Breaking-the-Google-reCAPTCHA)

5 算力验证码

验证码的初衷是人机识别。但在某些场景下,只需要降低用户频率，增加时间成本即可达到需求。

可以使用一种基于密码的，基于哈希的工作量证明算法Hashcash需要选择可选的工作量来进行计算。

将哈希戳的文本编码添加到请求的头中，以证明发件人在发送请求之前已花费了适度的CPU时间来计算戳记。换句话说，由于发件人已经花费了一定的时间来生成并发送请求，因此它们不太可能是垃圾发送者。接收方可以以可忽略的计算成本来验证该戳记是否有效。

该方法在电子邮件系统、论坛系统中应用广泛。

1. 优点：通过无差别的算力要求可以稳定提升攻击方的成本，对普通用户影响微乎其微。除了暴力穷举没有其他已知的破解方法。
2.
3. 缺点：主要应用于过滤大量垃圾信息的发送，对于有特定目的的攻击者可能不能提供一个很好的过滤。而如果过于提升算力要求，可能对低级硬件的用户造成使用上的负担。

补充：
目前该方法的一种破解思路在于：

利用本地计算的高性能，不使用服务器提供的js代码，而是使用更高性能的语言计算；甚至调用高性能的GPU计算，利用GPU在哈希计算上的优越性来提高计算效率、减少时间成本。

* 其他验证码

为能力缺陷人员设计的验证码，包括语音验证码、声纹验证码等。

本质和图片验证码没有区别，都是对信息进行扭曲、干扰、链接、变形以达到机器无法识别的效果。

不熟悉、不清楚、不了解。

二、部分验证码项目

1 google-reCAPTCHA

reCAPTCHA是一项由google提供的免费服务，使用风险分析技术来区分人和机器人。

其最新的reCAPTCHA v3允许网站在没有任何用户交互的情况下验证交互是否合法。它是一个纯JavaScript API，返回一个0到1之间分数来标记来标记合法的概率。

1. 优点：不会打断用户操作；返回一个可能概率，可以和其他验证码系统无缝组合；一个免费的验证码api，由专业组织维护、更新。
2.
3. 缺点：来自google，使用api而非开源项目。
4.
5. 文档链接：https://developers.google.com/recaptcha/docs/v3

补充：
vue-recaptcha

vue-captcha

2 kaptcha

由谷歌开源的验证码生成项目，可以将数字、图片进行扭曲、链接、加噪得到一个图片验证码。

1. 优点：成熟，方便配置。
2.

3. 缺点：比较老旧，安全性存疑；用户体验差。
4.
5. 文档链接：https://code.google.com/archive/p/kaptcha/

3 EasyCaptcha

github上基于java的图形验证码，支持gif、中文、算术等类型，可用于Java Web、JavaSE等项目。

1. 优点：支持种类较多，支持中文验证。
2.
3. 缺点：内置字体使用的种类、扭曲方式等依靠列表，容易被暴力遍历。
4.
5. 文档链接：https://github.com/whvcse/EasyCaptcha

4 /verify.js

托管于github的开源项目，提供常规验证码、滑动验证码、拼图验证码、选字验证码，纯前端验证码。

官网进不去。github仓库为https://github.com/Hibear/verify

5 SliderCaptcha

滑动验证码,用户通过滑动滑块来进行校验,可将用户拖动行为的时间、精度、轨迹等信息上传至服务器,然后进行后台算法验证。

1. 文档：gitee.com/LongbowEnterprise/SliderCaptcha