

一种基于常识的图片验证码设计方法

何春辉

(湘潭大学 工程训练中心 湖南 湘潭 411105)

摘要: 为了防止非法用户对在线系统的恶意攻击,提出了一种简单高效基于常识的图片验证码设计方法。通过不同难度层次的常识构建常识库,设计生成的图片验证码质量较高,用户根据图片验证码的常识问题推理得到答案,将答案作为密匙进行认证,而不是直接输入图片验证码上面的字符,可以有效杜绝非法用户对系统的攻击,弥补当前诸多验证方法的不足,同时又兼顾了验证码设计的可用性和安全性准则,适用于普通在线系统的身份认证。

关键词: 图片验证码; 常识库; 问答

中图分类号: TP393

文献标识码: A

文章编号: 1671-5322(2018)04-0042-03

随着互联网技术和爬虫技术的不断发展,很多B/S网络架构的在线系统受到了网络攻击,从而出现系统瘫痪导致无法为用户提供服务的现象。面对这种状况,仅仅使用用户名和密码的验证技术已经无法满足大多数网站的安全性要求^[1],而采用验证码认证技术可以有效解决大部分的网络攻击,为系统正常运转提供良好的安全屏障。因此,在系统登录阶段增加用户输入验证码的环节成为解决这类问题的简单有效途径^[2]。

目前市面上常见的验证码大部分是由随机数字、字母或汉字组合以及图片的形式生成,这种类型的验证码极易被破解^[3],导致系统的安全屏障失效。为了克服这种不足,文中提出了一种简单高效基于常识的图片验证码设计方法。该方法需要根据验证码的提示信息进行简单的推理,然后将推理得出的结果作为输入进行验证。这种方法可以有效提高系统的安全性,且容易操作,不会影响正常用户登录系统。

1 验证码

验证码一词最早出现于2002年,是一种区分用户属于计算机还是人的公共全自动程序,可以防止恶意破解密码、刷票等不良行为。

1.1 实现原理

验证码通常情况下都是由服务器端直接生

成,然后通过网络请求显示到前端的登录页面。浏览器端的用户根据前端登录页面上验证码提示信息填写表单,并提交到网站服务器端做身份验证^[4],若验证成功就可以顺利登录系统,否则需要重新做身份验证。实践表明,通过这种简单的策略可以有效提高系统的安全性。

1.2 验证码种类

随着技术的发展,目前验证码类型丰富多样且各具特色,比较流行的有如下6种类型:(1)数字验证码,一般是指验证码由4~6位纯数字随机构成;(2)字母验证码,通常是指验证码由4~6位纯英文字母随机构成;(3)汉字验证码,常指验证码由4~6个纯汉字或者汉字笔画^[5]随机构成^[6-9];(4)混合验证码,通常采用4~6位的数字、字母、汉字的随机组合来生成,它的安全性比上述三者更高;(5)图片验证码^[10-12],主要是指将指定位数的随机字符串转化为静态图片形式或者动态拼图的图片验证码,这种类型的验证码安全性非常高;(6)语音验证码^[13-14],是指将语音经过变调、混合背景噪声等处理后生成的验证码,这种验证码安全性高,但是相应的成本也较高,目前实际使用中并不多见。

综合考虑系统安全性和开发成本等问题,目前大多数在线系统都是采用图片验证码来做身份认证。

收稿日期: 2018-05-17

作者简介: 何春辉(1991—),男,湖南永州人,工程师,硕士,主要研究方向为算法研究与应用。

1.3 验证码设计准则

设计验证码的目的主要是强制进行人机交流,自动抵制非合法用户对系统的攻击,因此在设计验证码的过程中要采用让合法用户轻松认证,而非法用户很难认证成功的原则来进行。一套优秀的验证码设计方案应具备以下几个基本准则:(1)生成验证码的算法应带有随机性和动态扩展性,具备可用性和较高安全性的特点;(2)验证码的长度既不能太长也不能过短;(3)生成验证码的算法要比较完善,至少应该包含少量噪声;(4)生成的验证码和它对应的密匙一定要准确一致,从而保证认证质量。

2 常识预处理

人类知识库经过长期积累形成了众多有用的知识,文中把其中得到人类公认和众所周知的这部分基本知识定义为“常识”。这种类型的常识来源于不同领域、不同文化且为公众所周知,例如“太阳从东方升起”或者“ $1+1=2$ ”等等。为了有效运用这些常识来设计高质量的图片验证码,需要对这些常识进行筛选、整理和分类,从而构建可用的基本常识库。

2.1 常识筛选与整理

考虑到人类已有的常识种类繁多且数量庞大,需要根据某些准则来筛选部分可用的常识构建常识库。本文主要目的是筛选常识生成图片验证码,因此选用1.3节所述的验证码设计准则作为常识的筛选依据。

首先利用验证码的长度不能太长也不能太短的准则有效筛选出合适的常识;然后利用验证码的生成和它对应的密匙要准确一致的准则筛选出只有一个准确结果的基本常识。依据这两条准则可以对“中国首都是北京”这种类型的常识进行正确筛选并加入常识库,而对“刘翔获得奥运会的冠军”(长度太长)或者“人的常见血型(A/B/O/AB)”(答案多样)这两种与上述两条准则不一致的常识类型进行有效过滤。

根据上述筛选依据可以构建一个可用的常识库,然后将常识库里面的常识作相关的处理,确保常识是以问题和对应答案这种成对的形式出现。文中在构建常识库的时候约定常识中的问题在前,答案在后,问题和答案之间用短划线连接,例如“中国首都_北京”。这样可以保证后续生成的图片验证码和密匙一一对应。

2.2 常识分类

考虑到常识数量大、种类繁多,以及图片验证码需要具有高度的可用性和安全性。因此需要对筛选出来的常识进行分类,按照不同领域进行分类筛选,让不同领域的用户轻松认证身份。为了兼顾安全性,不让非法用户轻易破解验证码系统,需对整个常识库中的常识设置不同的难度等级,一般比例为8:2,即80%的基本常识,20%的一般难度的常识。采用这种设计比例,既可以达到可用性,又可以达到安全性。

3 基于常识的图片验证码设计流程

为了设计出既简单又安全可用的高质量图片验证码系统,利用2.1节和2.2节构建的常识库设计图片验证码,具体流程如图1所示。

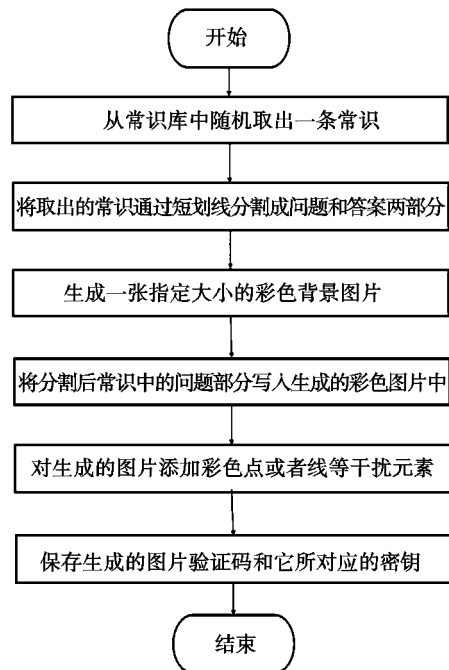


图1 基于常识的图片验证码设计流程

Fig.1 Image verification code design flow based on common knowledge

考虑到验证码的随机性准则,需要从常识库中随机取出一条尝试记录,然后将取出的常识通过分割标记符进行切分,得到常识的问题和对应的答案两个部分;接下来使用JAVA的画图工具包AWT生成一张指定大小(文中使用4个汉字的长度)的彩色图片,并将分割得到的常识的问题部分写入生成的彩色图片,再在彩色图片上添加一些彩色点或者线作为干扰信息;最后将得到的图片验证码和它对应的答案存储到指定的地

方,以便认证使用。

4 基于常识的图片验证码设计实例




按照图 1 的设计流程,通过简单示例最终得到部分图片验证码和它所对应的密匙,如表 1 所示。

5 结论

针对在线系统时常碰到客户机恶意攻击的行为,文中提出了一种简单高效基于常识的图片验证码设计方法。通过不同难度层次的常识构建常识库,设计生成的图片验证码质量较高,又兼顾了验证码设计的可用性和安全性准则,适用于普通在线系统的身份认证。用户根据图片验证码的常识问题推理得到答案,将答案作为密匙进行认证,

表 1 生成的部分图片验证码和对应的密匙

Table 1 Generated partial image verification code and corresponding key

生成的图片验证码	对应密匙
	赣
	昆明
	巴黎

而不是直接输入图片验证码上面的字符。采用这种问答形式的验证码可以有效杜绝非法用户对系统的攻击,弥补当前诸多验证方法的不足。

参考文献:

- [1] 龚莎. 网页验证码技术综述[J]. 信息与电脑(理论版) 2014(2): 143-144.
- [2] 陈义中, 李松林. 一种注册登录系统图片验证码设计与实现[J]. 软件导刊 2016, 15(7): 82-85.
- [3] 刘欢, 邵蔚元, 郭跃飞. 卷积神经网络在验证码识别上的应用与研究[J]. 计算机工程与应用 2016, 52(18): 1-7.
- [4] 方蓓. 互联网图片验证码的实现与验证[J]. 电脑知识与技术 2011, 7(22): 5301-5303.
- [5] 曹记东. 一种基于汉字笔画的验证码生成技术[J]. 重庆文理学院学报(社会科学版) 2016, 35(5): 98-100.
- [6] 常祖政, 卢晓倩. Java Web 中随机汉字扭曲验证码的实现[J]. 电脑编程技巧与维护 2010(11): 87-90.
- [7] 杨俊峰. 基于 Java 的中文验证码生成器的设计与实现[J]. 电子技术 2013, 42(9): 70-72.
- [8] 陈占芳, 冯欣, 张伟. 随机中文汉字验证码的生成及应用[J]. 电脑知识与技术 2007, 16(3): 1096-1097.
- [9] 戴如意, 艾丽蓉, 高海昌. 基于多形变特征的汉字验证码的设计及实现[J]. 计算机应用研究 2010, 27(4): 1508-1510.
- [10] 孔令芬. CPC 模式下图形验证码对恶意点击的有效预防[D]. 天津: 河北工业大学 2011.
- [11] 曹放. 基于视觉感知的图像验证码设计与研究[D]. 西安: 西安电子科技大学 2017.
- [12] 张守强. QR 码的数字水印算法和水印图片验证码的研究[D]. 郑州: 郑州大学 2017.
- [13] 郭峰. 汉语语音验证码技术及应用[D]. 杭州: 浙江大学 2010.
- [14] 汪成亮, 张玉维. 基于共振峰合成和韵律调整的语音验证码方法研究[J]. 计算机应用研究 2011, 28(7): 2458-2461.

An Image Verification Code Design Method Based on Common Knowledge

HE Chunhui

(Engineering Training Center, Xiangtan University, Xiangtan Hunan 411105, China)

Abstract: In order to prevent the malicious attack of illegal users on the online system, a simple and efficient design method of image verification code based on common sense is proposed. The common sense library is built by common sense of different difficulty levels, and the quality of the picture verification code generated by the design is high. The user infers the answer according to the common sense problem of the picture verification code, and authenticates the answer as a key instead of directly inputting the character above the picture verification code. It can effectively prevent illegal users from attacking the system and make up for the shortcomings of many current authentication methods. At the same time, it also takes into account the usability and security criteria of authentication code design. It is suitable for the authentication of ordinary online systems.

Keywords: Image Verification Code; Common Knowledge Library; Question and Answer

(责任编辑: 李华云)