SEMAGE:一种新的基于图像的双因子验证码

摘要

我们提出了SEMAGE（**Se**mantically **Ma**tching Ima**ge**s），一种基于图片的验证码，它利用人的能力去定义理解图片的内容和建立他们之间的*语义关系*。一个SEMAGE问题需要用户从给定的图片集中选择*语义关联*图片。SEMAGE有双因子设计，用户需要指出每个图片的内容然后理解和指出拥有语义联系的子集合。大多数的基于图片状态的图片系统像Assira[20]，只需要用户去解决第一个层次：图像识别。利用图像语义关联创建更加安全和用户友好的问题让SEMAGE新奇。SEMAGE并不会有像传统的基于图片的方法有的取法定制性和适配性的缺点。SEMAGE不像现在的基于文本的系统，它是很用户友好的很有乐趣的。这些特性让它对web服务提供者是十分有吸引力的。另外，SEMAGE是依赖语言的和对定制高度灵活的（在安全性和可用性层次）。SEMAGE还是移动设备友好的，因为它不需要用户输入任何东西。我们实施第一个这种类型的大伸展性的174个用户的研究来计量比较SEMAGE，reCAPTCHA（test-base）和Asirra（image-based）的精度和可用性。用户研究进一步验证我们的观点，结果表明用户使用我们的系统拥有更高得精度，并认为我们的系统是有乐趣的和简单的。

类别和主题描述

K.6.5[计算机空间]：计算机管理和信息系统 - 安全和保护

主要词汇

安全

关键字

验证码，基于语义的交互验证，双因子验证码

1. 前言

在生活的各个领域，每天都有新的web应用和服务出现。许多人都适应了在线服务，如email服务，论坛和专业兴趣组。对于服务提供者，一个重要方面是考虑怎样确保服务和资源能被目标客户使用。恶意使用服务，如使用机器人去注册合法账户[9]，不仅会占用宝贵资源，还未发布恶意信息埋下伏笔。所以服务提供商能从人类的访问中分辨出机器人是十分重要的，为了这个目的，验证码系统被广泛使用。 验证码（CAPTCHA）是“Completely Automated Public Tests to tell Computers and Humans Apart”[29,28,27,15,9]的缩写。主要思想是出一道困难的AI问题来区别当前服务的是机器人还是合法用户，除非AI技术获得突破性进展。验证码系

统的鲁棒性并非依赖数据库的安全性，而是这个问题固有的难度。解决验证码问题的难度对于人和机器人来说，通常有相同的难度。因为验证码系统很少是独立的，它一般是应用程序的一部分，如在线注册，让用户花多于几秒钟的时间去解决验证码问题是不现实的。于是，在现实系统中，一个复杂的问题需要人花费很长时间去解决是不现实的。 辨识扭曲的文字，基于图片的提问，是一种防止机器人的技术，使用非常广泛。然而，伴随着计算机视觉领域的发展，机器人已经可以破解使用如OCR（Optical Character Recognition）和语义[30, 26, 16, 2, 19]技术的文本验证码。引入噪声和扭曲来提高基于文本的系统复杂度虽然让机器人的破解变得困难，但是同样也降低了用户友好性和可用性。 基于图片的系统目的是提高验证码系统的可用性[20,3,17,23,18,7,25,32]。然而，许多现有的基于图像状态的图片系统如Asirra[20]有着灵活性和适应性不良的缺点。Asirra是难度只有图像识别，需要用户来找出所有的猫从猫和狗中。在像Asirra系统上，如Golle所示[22]，一种特别的机器学习技术的攻击方法已经获得很高的识别率。此外，呈现给机器人的固有选择总是二值（一个图片是猫或是狗），让它对模板适应攻击很敏感，这个问题将在4.2章中详细讨论。我们提出了SEMAGE，一种新奇的基于图像的验证码系统，它有双因子模型需要用户来识别图片和确定图片所共有的语义上的联系。语义上联系使得SEMAGE有更好的鲁棒性来面对相同的机器学习攻击。其他的基于图像的系统，像ESP*PIX[3] 和 SQ*PIX[7]有语言依赖和可使用性的担忧。我们将说明更多的验证码系统和它们的局限性在2章中。 在本论文中，我们提出了SEMAGE（**Se**mantically **Ma**tching Ima**ge**s），一种双因子验证码系统。在SEMAGE系统中，我们把一系列的候选图片呈现给用户，而不呈现拥有语义联系的子集合。对于用户的挑战是在内容中确定系统定义的语义相关的图片。需要注意的事，在正确集合里的图片不需要拥有相同的对象，一个语义相关的图片集合可能是不同物理属性单却有相同的定义内容的实体的图片组成的。考虑这样一个例子，用户需要回答内容相似拥有相同起源的图片，候选集包括图片如木柴，木椅，火柴，电器，一个动物和一个人，木柴，木椅，火柴属于一个相似集。 SEMAGE验证码系统难度体现在两个方面：（1）用户需要独立的指出每张图片的内容，也就是图像识别，（2）还需要在理解图像的语义关系的基础上正确的匹配文件。人类会很自然的采用他们的认知能力和平常直觉来解决这个问题，甚至不会感觉到这个问题固有的难度。同样，机器人也要面对相同的难度，即理解图片并指出它们之间的关系，这构成了一个很难的AI问题。我们的双因子设计主要目的是提高机器人破解的难度等级的同时提高用户友好性，而不牺牲系统的鲁棒性。SEMAGE的新奇之处是呈现给用户双因子问题：“在给定的内容中确定相同语义的图片”。展示语义相似的图片让用户选择的点子比简单的选择相同物种的动物图片更加宽广（Assira中的例子是猫）。这个特性让SEMAGE与基于图像状态的图片不同，它仅仅需要用户解决第一个层次的问题：图像识别。计算机很难去理解并确定图片内容的语义，这使得SEMAGE对于机器人有很好的鲁棒性。我们将陈述和讨论一些相似语义的需要在3章。 我们也实现了一个非常简单的SEMAGE例子，使用真实的和卡通的动物图片。用户需要挑出相同物种的图片（真实的或是卡通的）。这种特别的实现会有明显的好处：（1）因为用户可以很容易的在真实和卡通图片之间建立连接，所以没有给用户增加负担，却增加了用户的乐趣；（2）因为机器人卡通图片可能没有真实的物理上的动物的属性，从而增加机器人破解的难度等级。除此之外，SEMAGE提供了一个容易操作的接口来指示正确的答案，这使得它对输入比较困难的设备如基于触控的系统和小的手机来说是理想的选择。图片1是一个简单的简化了的SEMAGE。人们可以确定图片的标签，而机器人却很难找出真实和卡通图片之间的联系，因为形状和纹理是不同的。需要注意的是，这里只是一种SEMAGE的实现。其他任何的*语义关系*都可以被用来当实体，来替代我们的简化的实现。 这篇论文的主要贡献如下：

- 我们提出了SEMAGE，一种基于图片的双因子验证码，它有价格独一无二的特性。SEMAGE的设计使用简单的办法提升了自然的和受欢迎的网站的安全性和可用性。SEMAGE验证码图片可以变化，使其适应不同需求的网站。事实上，给图片标示上*语义联系*是很简单，可以根据直觉完成的事。我们提供了深层次的安全分析来展示SEMAGE对于许多攻击有更好的鲁棒性比起现有的系统。
- 我们更进一步的使用简单的简化的SEMAGE实现，实施了一个大尺度的用户研究，174个用户参加了这

个研究。我们系统与基于文本的验证码系统reCAPTCHA[6]和基于图像的Asirra[20]在可用性和娱乐性上进行对比。正如5章节讨论的一样，结果显示我们的系统更加容易去使用，参与者对我们系统的娱乐性给出了很高的评价

2.背景

CAPTCHA系统，特指基于文本的，作为反机器人首选方案，已经被广泛的使用在互联网中。最近，随着计算机视觉技术的提高，基于文本的系统变得对于机器人的攻击易感，攻击拥有很高的成功几率[30,26,16,2,19,13]。因此，很多精力花在了如何替换CAPTCHA系统，如基于图片的[20,3,17,23,18,7,25,32]和基于声音的[14,10,1,21]系统。

2.1 基于文本的系统

大体来说，基于文本的验证码系统让用户辨识字母或者数字。GIMPY是一种经典的例子[4]，攻击基于文本的系统大多使用OCR(optical character recognition)算法。这个算法首先把图片分割成小的区块，每个区块只有一个字母，然后使用模式识别算法去匹配每一个区块和字母模板的特性[30,26,16]。这里的后一步是一个很成熟的AI问题。为了防止这样的攻击，基于文本的系统使用了如下的技术来加强鲁棒性[15,19]：

- 添加一些噪音，如一些散落的横线和点，到背景中来干扰区域分割算法。
- 字符被连接起来，或者互相重叠起来，来让攻击算法无法正确的划分图片到正确的区块里。
- 字符是扭曲的来增加文字识别的难度。

然而，以上的技术提高了人类识别的难度。字符的连接让人类也很难识别。举个例子，让字符‘r’和‘n’被连接起来，就像是字符‘m’。扭曲的字符不仅折磨用户的神经，还会影响准确率。图2展示了一个很难去解决的基于文本的问题。 基于文本的系统面对一个必然的处境：当验证码系统变得复杂后，人们会解决觉得验证码很难受的。这可能是一些热门的网站如MSN邮件使用简单的干净的的验证码，但是破解的准确率高于80% [30]。一些系统使用特别的颜色来标示每个字符，再给背景添加非字符颜色。然而这些都能很容易被自动化的程序移除，并没有个机器人添加任何的困难[31]。 热门的系统，如‘reCAPTCHA’[6]使用词典单词，这些单词是被真实的自动化OCR程序标示为无法识别的，这个过程也被来数字化图书和验证其他用户数据如的正确性。然而，reCAPTCHA同样也减少了用户友好性和用户满意度，因为一些高度扭曲和噪音。

2.2 基于声音的系统

基于声音的验证码系统[1,14,10,21]补偿了可视化验证码系统无法满足视觉障碍的人的可用性需求。在一个典型的基于声音的系统里，字母和数字被随机的时间间隔隔开来以发音的方式呈现出来。为了使它对机器人更加有鲁棒性，背景噪音被添加到音频文件中。这样的系统很大的依赖声音硬件，并且用户只有很少的时间去确定每一个字符。一些人认为，声音验证码系统仅仅是基于文本的系统的听觉版本。虽然可视化的东西呗声音代替，但是构成攻击的基础都是相似的 - 特征提取和分类字母。对于机器人和人类的难度曲线是相似的[12]。所以声音验证码系统既没有提供更加用户友好的接口对于视觉访问用户，也没对机器人有更好的鲁棒性[11]。

2.3 基于图像的系统

基于图像的系统的出现时为了替代对于人来说越来越复杂的基于文本的验证码系统，基于图像的系统人类更好去解决。设计一个好的验证码系统，安全性不仅仅是考虑因素。所有的验证码系统都是HIP（Human Interactional Proofs）的一种形式，需要用户加入。所以用户友好性在设计中是一个很重要的因素。Tygar[17]提出了一个好的验证码系统有如下需求：

- 对于人类来说，任务应该是很简单的
- 对于计算机算法来说，任务应该是很难得
- 数据库应该能很简单的被实现和分析。

基于图片的验证码主要利用图片比起文本有更多信息。对于人类来说依靠直觉能很好抓住图片的额特征而对于AI算法来说，却是很难得。ESP-PIX[3]是呈现给用户一系列的图片，然后需要用户从一些类词语中选择一个能描述所有图片的。这中方法有两种缺点：它还是依赖文本去表达含义，因为所有的单词都是英文，用户验证成功很大程度上依赖他对英语的熟练程度（或者它移植到的其他语言）。它不仅仅依赖语言，还很难去实施。用户需要大致浏览所有的给出的单词，然后找到一个最可能的单词。SQ-PIX[7]也给用户一系列图片，但是让用户选择一张定出物品名字对应的图片，并标示出物品所在位置。这同样也是依赖于语言的而且在指定图片的时候需要一个手持设备如鼠标来操作，并不能对所有用户都简单。 Google的图片验证码‘what's up’[23]需要用户去调整图片的方向。这样的系统是语言无关的，但是调整的过程需要很多的精力和微小的鼠标（或者其他硬件）移动。还有一些图片可能有些模棱两可因为它有多个正确的方向。 Microsoft的Asirra[20]利用了在petfinder.com上的已有的数据库，呈现给用户许多猫和狗的照片，然后让用户在12个宠物中确定所有猫的图片。这个平台是语言无关的，需要用户查看12张图片，然后平均点击6次。图3显示Assira的一个简单例子。 Asirra配合petfinder.com可以访问他的很大的有猫和狗的数据库。但是固有的难度对于机器人来说仅仅是区分猫和狗。这使得Assira是不稳固面对机器学习攻击[22]。而SEMAGE在有双因子设计，用户需要识别每张图片，还需要理解并确定它的子集语义上的联系。而Assira只需要用户去解决第一个层次（图像识别）。利用图片之间的语义相关性，制造更加安全和用户友好的测试让SEMAGE有更好的鲁棒性。

3. SEMAGE设计

我们提出了SEMAGE，“**Se**mantically **Ma**tching Ima**ge**s”，一种新奇的基于图片的双营子验证码系统，他利用图片之间的语义关系。语义上的查询已经被用在其他领域如web搜索[24]。我们构想定义语义相似的图片然后设计一个利用这样概念用户友好的和有更好鲁棒性的验证码系统。

3.1 直觉思想

所有的基于图片的验证码系统有两个组件：一个图片数据库和一个怎样去使用数据库创建问题的‘概念’。固有的概念和PIX[8]一样简单，就是从数据库中抽取一系列有相同物体的图片展示出来，然后让用户指定合适的图片标签或者更复杂的东西像Cortcha[32]。Cortcha使用书库创建一个待修复的图片和一些候选图片，然后让用户把正确的图片放置到待修复的图片中。 SEMAGE背后的思想是使用图片之间的语义上的关联，让用户选出语义相关或是相似的图片。语义关系是相似图片背后的真实的描述。语义关系的选择对于一个应用来说有很大的自由度，数据库给了很大可定制的灵活性。举个例子，对于一个电子商务网站，SEMAGE可以以产品图片形式显示出来（ipod，zune，电视，燃气热水器，冰箱等等），让用户选择能干相同事的东西（在这个例子中是ipod和zune，都是音乐播放器）。 SEMAGE呈现一系列的候选图片，这些候选图片的子集有相同的隐含的连接或是关系。对于用户的问题是正确的确定所有图片的语义上的关联。

3.2 定义语义关系

我们现在陈述选择‘语义相似’关系的图片来创建验证码问题。‘语义标签’是一个专业词汇表述一个对象的有关系的标签。语义标签可以被直接的使用来标示数据库创建问题。让$SL(x)$标示返回x的语义标签的函数，我们认为两张图片是语义相关的，如果他满足下面的任意一种情况

- 情况1：如果两张图片拥有相同的语义标签。给出两张图片A和B，他们如果$SL(a) = SL(b)$，则认为他们有语义上的联系。例如，一个电脑的图片和一个电视的图片被定义一样的语义标签(SL)‘电器’
- 情况2：两张图片被分类到相同类别的语义标签。给出两个图片A和B，他们有语义上的联系如果$\exists\ T\ s.t.\ SL(A) \subset T\ \&\ SL(B) \subset T$，其中$T$代表一个语义标签。举个例子，一个狮子的图片和一个鹿的图片可以被分到相同的语义标签下‘四条腿的动物’。相似的，一个电视的图片和一个电脑的图片可以被分到相同的语义标签‘电器’。
- 情况3：当两张图片放到一起的时候，他们展示了独一无二的可辨识的概念。给出两张图片A，B和一些语义标签C，其中C代表需求的集合，A和B是语义上合适的如果$\{A \cup B\} \vDash C$其中⊨标示左边的满足右边的需求，举个例子，一张打印机的图片和一张纸的图片也可以被定义为可以辨识的概念‘打印’，这是一个语义标签。

‘语义关系’的需求更加普遍，语义相关增加，因为我们可以将情况1移动到情况3.为了建立一个SEMAGE问题，我们应该确保只有一个图片的子集可以满足上面的情况，最好给图片最少的广义标签。

3.3 问题创建

我们开发了一个简单的逻辑去创建SEMAGE问题。首先我们定义需要的相关参数如下。 n是一个问题中图片的数量，m是相似的或是有关系的图片数量。U是一个超集，代表数据库中所有图片的集合。每个测试的集合表示为S，也就是$|S| = n$。存在的‘语义相关’的子集为R,让所有的在R中图片有相同的语义标签。举个例子：$\forall\ r_i, r_j \in R, SL(r_i) = SL(r_j) \& |R| = m$，$D$是图片的集合，$|D| = n - m$，$D$中的每一张图片有与$R$中元素不同的语义标签。也就是$\forall\ d_i, D_j \in D, SL(d_i) \neq SL(d_j) \neq SL(R)$。这表示所有在子集$D$中的元素和$R$中的元素有不同的语义标签，这样可以使$R$中的有语义关联的图片不是模棱两可的。这样的话，每一个问题的集合$S = R \cup D$。 我们现在描述一个简单的算法来实现验证码问题生成，详见算法1.数据库由有语义标签的图片组成。算法开始的时候R和D都是空的。我们随机的从数据库中选择出标签，然后用有相同标签的图片填充R。然后我们用有和先前D和R中选择了的图片不同标签的图片填充D。R和D中图片数量由n和m决定，这个是由用户指定的。R和D中的图片被以随机的顺序以表格形式呈现给用户。

3.4 数据库

填充数据库对于基于图片的同时是一个重要的问题。不像文本验证码，可以使用随机的字符组合来创建验证码问题，SEMAGE中的图片需要语义相似，这个需要小心的被设置。一些人可能随意使用搜索服务（像Google）来搜索相关的图片。而在我们的实现中，我们开发一种半自动化的机制在爬Internet时填充数据库。也同样可以从电影中或是短片中提取。以上所有方式都是半自动化的，需要一些人力清除不相干的图片。这种方法的缺点是攻击者可以冒险的花费足够多的时间和人力来重新构建整个数据库。 因为SEMAGE固有的设计，它提供一个为网站创建数据库的方法，像电子商务网站，已近有了一个图片数据库。电子商务的web提供商一般同一个商品的多张图片（如从不同角度的图片），同一种产品的不同样式（相同的产品，不同的颜色，尺寸，包装），多重相同类比的商品。图片被产品信息标示，产品信息可以分到不同的类别。使用‘内容

语义’多种关系可以被建立。有了富足的存在标签的信息，我们只要添加很少的逻辑就可以实现‘验证码问题创建’算法。此外，一些数据库还实现了复杂的关系，如‘相似的产品’，当用户浏览某件商品的时候可以推荐类似的商品。这样，更加复杂的‘语义关系’可以利用这样的信息被建立。使用这样的图片不仅仅可以让数据库更加安全，还可以当做商品的广告。

4. SEMAGE分析

4.1 设计分析

4.1.1 可用性

安全可用性是SEMAGE重要的关注点。图片包含的内容对用户的认知是有意义的，是很容易辨认的出的。利用人类广阔的常识存储，我们的设计可以让用户花费很少的努力来解决问题。此外，它合乎人的思考方式 - 人类看到图片的第一眼就是看看图片关于什么，而不是考虑那些细节（方向，某些图片的特性）。建立物体之间的联系人类很自然，人类也可以自动的去掉那些模糊不清的东西。举一个例子，一个红色骑车被呈现在其他颜色的车里，人类立刻就能注意颜色的不同。然而，如果相同的红色汽车和红色水桶，红色衣服呈现在一起，人类立刻就能注意到物体类别的不同。对于电脑来说，这里的每一步都是很难的AI问题。在第一步需要进行图像识别，去确定图片包含了什么，并标上预标签。解决‘关系’问题，计算机不仅仅需要很大的正确标签数据库，还需要很复杂的AI直觉。这样对人和电脑来说，有很大的难度差异。 另外，SEMAGE对用户来说，提供了一个简单实现的接口去标出正确的答案。只需要鼠标简单的点击即可选择正确的答案，这使得SEMAGE对触屏系统和小的电话来说是一个理想的选择，因为打字是和困难的。这也比跟踪突出的物体简单得多（如SQ-PIX[7]）。

4.1.2 语言依赖性

我们的设计利用图片语言的约束。一些验证码系统也使用语义线索如ESP-PIX[3]。然而它需要用户去寻找正确的单词，在一系列的描述图片内容的单词中。这就受限于用户对语言的熟练程度。我们的税基是语言无关的，可以被全世界的用户使用。这特别的使不习惯把英文当做日常语言的人。

4.1.3 定制灵活性

我们的设计提供了几种方法去自定义问题，包括内容，安全等级，可使用性等级。图片数据库可以被定制来适应需求和宿主网站的样式。如，对于特定的利益集团，数据库可以是集团的主题的物品，如电影截图之于电影出租网站或者特殊产品之于电子商务网站。这提供了可能的内容广告或者提供了一些乐趣，而不是传统的无聊的验证码测试。 它也同样很容易的定制安全等级对于网站的管理员。管理员可以决定图片池的大小和正确答案集合的大小。考虑这样的设定，n是候选图片的数量，用户需要选择k张匹配的图片，随机猜测成功概率为$1/C_k^n$。提高答案集合的大小并不是很必要来减小随机猜测的成功概率当n很小的时候，但是当n增加的时候，随机猜测攻击的成功率下降。考虑用户体验，用户使用在验证码上的时间随着候选图片池中的图片数量提升而提升，但是提高答案集中图片的数量并不明显增加用户使用的时间。我们认为选择最佳的n和k决定于实际使用的图片的内容，如果需要这样的数据，可以进行一个特别的用户研究。

4.2 安全性分析

我们考虑一个对手模型，机器人可以访问一个没有标签和分类的图片数据库，而这个数据库就是我们的出题来源。这里需要注意的是，如果给足够的时间和资源，下面讨论的攻击可能能成功，但是需要花费长时间类破解，也是这类系统设计的最初的目的。我们的目标正如其他的验证码系统一样，让当前的攻击尽可能的困难，这样任何成功的攻击需要在技术上有大幅度的进步。我们现在确定和分析一些可能的攻击方法来攻击我们的系统和需要花费多少来避免他们。

4.2.1 使用机器学习的方法来攻击

相似的技术被使用来攻击Asirra[22]也同样可能被用来攻击我们的系统。攻击Assira的方法可以用来攻击我们模型的第一个层次'图像识别'，本质来说，攻击者尝试获得确定数量的正确的有标签的图片，用来训练几种不同的分类器，要么是基于色彩信息的，要么是基于纹理信息。然而，解决一个SEMAGE问题不仅仅需要图像识别，更需要确定'语义关系'。确定图片之间的'语义关系'是一个还没有解决的AI问题。更进一步，即使语义联系是很弱的，语义标签就是物品的名字，SEMAGE可以容纳更多的物品类型比起Asirra（仅仅为2），所以攻击者需要构建可以进行更多类型分类的分类器。 现在让我们考虑有一个非常简单的'语义关系'，相同动物的'真实和卡通'图片（如5章中使用的）。卡通和真实的动物的物种的颜色和纹理数据变化要比卡通和真实的动物大得多，正如图4说明的。然而，攻击者可以尝试训练分类器独立的分类真实动物和卡通动物，性能将会随着分类器数量的提升而降低，而且会变得很复杂。所以使用这种算法攻击的成功率将会很低。

使用模板适应技术的攻击：在图像识别领域，一个开发区域是适应物体到特征的模板上。举个例子，一个椅子可以被确定，如果给一个模板'四只脚，水平的上顶'。因此，对于一个兔子来说，特征可能是'朝上的长耳朵'。然而，这是很难定义'长的'比起'朝上'。一个鹿有一个朝上的耳朵可能被分到'兔子'模板。更进一步，不是所有的物体都有这样独一无二的确定的简单的特征

4.2.2 随机猜测攻击

对于一个SEMAGE模式陈列n个候选图片需要用户选择k个匹配的图片。成功随机猜测概率为$1/C_n^k$。正如图5所示，选择一个小的n和k可能让系统对随机猜测攻击脆弱。另外一方面，一个小的n，k是系统有 更好的用户友好性，让用户更少遭受挫折。我们实现的为了用户研究的系统使用了较小的n，k值，这使得系统对随机猜测攻击更加敏感。在小的n,k系统中，多个轮回的SEMAGE可以构成一个问题。这样的技术已近被使用在现有的系统如reCAPCTHA。选择相关的小的n，k值，我们牺牲了一点安全性来换来可用性。我们这样做是因为我们可以弥补SEMAGE对于随机猜测攻击的敏感，我们可以使用令牌桶（Token Buckets）[20]系统来阻止暴力攻击者。Assira需要更多的图片在每一个问题集被保护因为物体只有有限的鉴别集合（实际只有2，只有猫和狗），而从理论上来说，在我们的SEMAGE实现中，有上千种不同的类别。这样SEMAGE的双因子设计允许我们使用小的n，k而不太多的牺牲系统的安全性。 一个SEMAGE系统同业也可以被其他技术实现，如部分信用算法（Partial Credit Algorithm ）[20]，它允许一个比较大的n，k，然后定义一个'大体上正确的'回答集合，其中缺失一张图片。令牌桶（Token buckets）[20]同样可以实现防止暴力攻击者进行连续的随机猜测攻击。

4.2.3 使用静态图片名字资源来攻击

如果HTML标签的源代码中含有图片的名字，攻击者可能潜在的使用这些名字来确定相似的图片。然而，这

类的攻击可以被在源代码中使用随机名字的方法很容易的防止。在我们的系实现中，图片的名字不会暴漏给了用户，THML中的图片名字被随机化，然后发给用户的。

4.2.4 使用根据系统使用关系创建的攻击数据库

攻击者可能手动定义全部的‘语义关系’，然后搜索并构建图片的仓库，从而建立攻击仓库。使用有标记的图片的数据库，暴力搜索来破解候选集合可能让他获得‘相似的’正确的集合。但是匹配每个问题中的图片可能花费很多的时间和资源，而不能构成可行的攻击；也有可能超出每个问题的最大提交时间。

4.2.5 使用挖掘图片的纹理信息来攻击

潜在的攻击者可能使用像Google的goggle系统，一个基于图片的搜索系统，来发现候选图片集的纹理描述，来确定图片之间的而联系。我首先讨论图片识别或是搜索现在仍然不够成熟（对于未知图片，仍然是一个很难的问题）。另外，确定图片之间关系，仅仅用纹理描述，仍然是一个很复杂的未解决的AI问题，特别是正确的相似图片决定于语义内容。这样的攻击潜在可能攻击大多数基于图片的系统，如Assira，PIX，SQ-PIX，但是因为SEMAGE的双因子设计，机器人任然需要理解和定义语义上的联系。纹理描述可能只能解决图片识别问题。可能存在一些图片在描述上有重叠，但是却不在同一个‘语义相似’的集合上。考虑这样一个例子，候选集合是‘四条腿’的动物。其中包括昆虫，鹿，柿子，人类，电子器件和其他不相关的东西。即使能对比纹理描述，但是对于机器人来说，确定狮子和鹿的关系还是很难。

5. 评估

我进行了一个大型的用户研究来评估SEMAGE的可用性，与Assira和reCAPTCHA做对比。为了这个目的，我们首先创建了一个网站，可以提供给用户简单的SEMAGE问题。

5.1 SEMAGE的简单实现

在我们的简单实现中，每一个问题由确定数量的图片组成（图片的数量是可以配置的），图片集合的一个子集中的图片有相同的唯一的关系或是特征。图片更进一步的进行随机的噪声和简单替换纹理的处理。我们的实现使用了PHP实现和MySQL当做数据库。图6是高层的实现设计。

选择‘语义关系’：在我们的实现中，问题集合由真实的和卡通的动物图片组成，关系子集的定义相同动物的‘真实和卡通的图片’。使用‘真实的和卡通的’关系去定义图片的‘语义关系’有以下优点：

- 相同动物的真实的和卡通的图片的关系是微妙的和多种多样的。原因是真实的和卡通的动物的可视化属性可能完全不同如形状，大小，轮廓
- 人类有固有的能力建立不相似的对象的可视化属性，这样可以很容易的通过测试，而当前图片状态的机器人却不能。我们将测试这样假设在用户研究中，这个将在5章中详细讨论
- 生成一个大的数据库是很容易的。一个简单的动物搜索，即可从iamges.google.com中获得成千上万的实体。因此，我们可以很简单的很快的建立一个很大的数据库。

图7显示了一个简单的我们实现的SEMAGE测试。我们的问题中图片的总数为6，‘语义相似’的图片为2张，其中一张是真实的图片，而另外一个张是相同动物的卡通图片。

数据库生成：SEMAGE实现的第一步，在‘相似’图片上定义语义联系，是数据库生成。在图6中展示了一个实现了的图片搜索和下载工具当做图片接收器，其中搜索和下载需要WEB上图片。这个工具接受搜索关键词（来搜索真实的和卡通的动物图片），图片维度，图片下载数量和标签。它将会自动的下载图片，并存储在数据库中。一个简单的动物搜索，即可从iamges.google.com中获得成千上万的实体。因此，我们可以很简单的很快的建立一个很大的数据库。实际中个，因为自动匹配不总是能找出相关的结果，我们需要从集合中手动剔除无关系的图片。

动态噪声添加：为了让基于机器学习图片分类器的攻击变得困难，我们在每一个问题的创建过程中随机的在图片中添加一些噪声。我们填在噪声在随机的形状里，再进行色彩的替换在ImageMagick库的帮助下[5]。添加的噪声形状可能是图片的中心也有可能是图片的边缘。我们同样也做一些颜色调整来防止机器人分类器简单的将噪声移除。这样的而随机噪声策略确保了每一张图片有不同的噪声等级。图8显示了SEMAGE问题添加噪声后的样子。

接口：正如图8和图7显示的一样，每一个问题由一系列图片组成。表格的标题描述用户需要点击的相似图片，然后点击提交发送响应到服务器验证。我们实验了不同的布局，如图片每个之间都距离很远，或者是图片是在一行或是一列，最后我发现图片在一起形成一个表格可以更加简单的确定相似的图片。

5.2 用户研究方法论

一个综合的IRB用户研究方法被实施来收集SEMAGE到底有多么的用户友好性的数据,这是一个被部署在实际系统上的验证码本质标准。我们结合了reCAPTCHA，一种基于文本的系统，和Asirra，一种基于图片的来自微软的系统，在用户研究中比较分析。Assira和reCAPTCHA在开放的web服务使用很多，很容易被我们的研究集成进来。志愿者远程参与这项研究，他们被给了一张纸的画报来告诉他们需要做什么来通过系统测试。我们记录他们完成每个问题的时间。用户可以得到他们是否通过了测试，在下一个测试出现之前。

一共174个志愿者参加了这个研究，主要人员是毕业学生和在校学生。用户池是各种各样的，大多数的用户都是没有计算机科学学科技能，还有一些英语为母语，或不是母语的人。用户池里有66个女性108和男性。所有测试的人都不知道我们的系统使用了SEMAGE。正如之前说明的一样，我们收集每一个用户完成测试的时间。我们也监控了所有尝试不论是否验证通过。我们也同样收集每一个测试成果和失败的次数。

5.3 用户研究结构

用户研究通过网页，分为以下几个阶段

- 一个初始化的问卷需要回答他们对验证码的熟悉程度，英文的熟练程度和其他人口统计学的问题如性别和年龄区间。
- 一张纸的潜在描述SEMAGE，Asssira和reCAPTCHA，让榕湖知道怎样去解决每一个问题。
- 5个不同的SEMAGE问题
- 5个不同的Asirra问题
- 5个不同的reCAPTCHA问题
- 一个最终的简短的问卷问用户SEMAGE与Assira验证码的娱乐程度和易使用程度。 我们相信对于每个系统画报般的描述是很必要的对于公平的使用性数据统计在图像识别系统中。用户可能是第一次看到基于图片的验证码，因为用户在过去可能总是使用基于文本的问题。向他们展示每个系统怎样能通过测试，

可以让我们收集更加公平的数据。这个研究平均花费8.7分钟完成。 我们把可用性评估分为以下几个维度：

- 用户完成问题的速度
- 用户需要多少次尝试才能通过测试
- 用户认为系统是有乐趣和简单的吗?

5.4 时间统计

正如表1显示，用户完成基于文本和SEMAGE问题速度快于Asirra。几乎所有的用户需要6秒钟以上完成Assira测试。

	SEMAGE	**Asirra**	**reCAPTCHA**
Time Taken in seconds	11.64	17.35	11.05

图9是时间分布图表，图中显示出每一个SEMAGE问题使用11.647秒或者更少，相反Asirra用户使用的时间在17.355秒左右。大部分的数据都是一致性均匀的，平均时间并不很大程度上收那些离散点的影响。这代表了大部分用户的行为。

我们发现，用户解决SEMAGE的时间与reCAPTCHA相似。这真的很令我们惊讶。我们原以为解决SEMAGE会比reCAPTCHA速度慢得多，因为基于文本的验证码被用户查分时间的使用，用户已经熟悉他们了，而我们的系统很多用户都是第一次看。这个令人振奋的结果表明，SEMAGE是一个非常用户友好的，而且比较容易去使用的。 我们承认Assira问题由更多的图片组成，可能需要使用更多时间去解决。然而，Assira每个问题需要更多的图片来确保安全，因为只有很有限不同类别的主体（只有两种，猫和狗），而我们的SEMAGE系统从理论上来说，有成千上万种不同的类别。更进一步来说，如果只有两个类别，用户只需要将对象放到两个类别其中一个类别里。另一方面是SEMAGE需要用户联系两个或者更多的图片，这可能需要用户花费更多的时间。然而，时间数据显示，SEMAGE问题比它更简单，因为人类的自然认知能力。

5.5 精度数据

简单的说，SEAMGE正确的的尝试比Asirra高。这显示了用户可以更加正确解决更多SEMAGE问题。图10(a)展示了SEMAGE和Assira正确的尝试的数量。在问卷的一开始我们同样问用户他们对验证码的熟练程度和适应程度，从1到5打分（5是最好的），正如我们在图10中看到的一样，即使是参与者他们自己对验证码系统熟练程度打分较低的（小于等于3），他们依然能表现出很好的正确率在SEMAGE和reCAPTCHA上比起Asirra。

为了让系统能被部署在真实的场景中，需要对人类有一个很高的尝试正确率。‘尝试正确率’（C.A.R）是正确的尝试除以所有尝试。这意味着人类需要多少次才能通过测试。这个数字越接近1，系统的可用性越高。

用户研究数据表明我们系统（0.94）比起Asirra（0.91）更高。用户对基于文本验证码系统更加熟悉，我们期望他们可以在reCAPTCHA系统中干的更好。但是再次，SEMAGE比起传统的基于文本的系统，完全差不多。这数据表明，我们的系统有很高的可用性比起现在基于图片状态的图片系统（Asirra）。

5.6 娱乐性和简单使用性

在对比三个系统之后，用户被问及比较SEMAGE和Assira的娱乐性和轻松性。这里有两个独立的问题：之一是娱乐性，另外一个是轻松性，让用户选择如下的打分：

- 1，如果用户认为Assira有更好的娱乐性或轻松性
- 3，如果用户认为Assira和SEAMGE有相等的娱乐性或是轻松性
- 5，如果用户认为SEMAGE有更好的娱乐性或轻松性
- 2或4，如果用户倾向于Assira或是SEAMGE

这些因素给我们提供了更加主观的指示器。我们可以很清楚的从图11中看出，大部分用户（50.92%）选择4或者5，这表明SEMAGE有更高的乐趣。仅仅16.07%的用户选择1和2，这表明Assira有更高的乐趣。这很明显的支持更多用户发现解决SEMAGE系统问题有更多的乐趣，比起Assira。

图11显示，简单使用性的打分分布。72.61%的用户选择了4和5分。只有10.72%的用户选择1和2分。16.66%的用户给出3分。

这些尺度和之前的时间，正确率结果清楚的证明SEMAGE是一个更高的用户友好的验证码系统。

6 局限性和以后的工作

生成一个巨大的正确的数据库对基于图片的系统还是一个很大的挑战。在我们的简单SEMAGE实现中，我们爬网页，自动收集和给图片添加标签。然而，不是所有的被爬虫返回的图片都是有关联的，一些甚至会让人反感。这样的手动劳动浪费了很多时间，还有可能会引起很大问题当数据库日常更新的时候。直接使用爬来的图片这里还有法律问题。

SEMAGE收益于他的设计思想，不需要用那种方法建立数据库。像电子商务，电影租赁网站可以很方便的使用已有的图片数据库，而这些数据已经有‘语义关联’。然而，建立自动的大的，正确的数据库还需要更多的工作。

这篇论文我们介绍了使用对象之间的‘语义关系’建立的验证码系统，然后实现了实现了另一个简单系统。我们实现的简单的系统并没有完全发掘SEMAGE的潜力，我们打算构建有更好鲁棒性的系统，更高层次以语义相关为基础的SEAMGE系统。

7 总结

这篇论文中，我们提出了SEMAGE（**Se**mantically **Ma**tching Ima**ge**s）。这种验证码系统呈现给用户一系列的图片，然后让用户选择语义相关的图片。这个问题分为两层：理解语义含义，建立语义关联。这个问题对用户来说是很自然的，因为它结合了轻量级的视觉和识别工作。然而，这种分层的结构，对反对机器人来说，提供双重的保护。因为交互接口是简单的和有效的，这是很容易理解的。验证码系统不断的寻找可用性和安全性之间的最佳的平衡点。SEMAGE给网站管理用提供极大的定制空间。在问题中，他们可以根据实际的网站对可用性和安全性的需要，定制候选图片和语义关联的图片数量。更进一步，SEMAGE可以和触控设备或是小的手机，这种输入比较困难的设备结合。网站管理员可以决定图片数据库的内容来迎合他们的促销

需求。SEMAGE数据库可以被特别的填充，或者从已有的数据库中适配过来。电子商务是一个SEMAGE数据库可以很容易建立的领域，SEMAGE同样也给可以被定制来满足安全性和广告宣传的目的。

8 参考文献

[1] Audio and visual captcha.http://www.nswardh.com/shout/.

[2] Breaking text captcha. http://www.blackhat-seo.com/2008/how-to-break-captchas/.

[3] Esp-pix. http://server251.theory.cs.cmu.edu/cgi-bin/esp-pix/esp-pix.

[4] Gimpy project.http://www.captcha.net/captchas/gimpy/.

[5] Imagemagick.http://www.imagemagick.org/script/index.php.

[6] recaptcha official site. reCaptchaOfficialSite:http://www.google.com/reCAPTCHA.

[7] Sq-pix. http://server251.theory.cs.cmu.edu/cgi-bin/sq-pix.

[8] L. v. Ahn. Human Computation. Ph. d. dissertation, Carnegie Mellon University, 2005.

[9] H. S. Baird and K. Popat. Human interactive proofs and document image analysis. In Proceedings of the 5th International Workshop on Document Analysis Systems V, DAS ’02, pages 507–518, London, UK, 2002. Springer-Verlag.

[10] J. P. Bigham and A. C. Cavender. Evaluating existing audio captchas and an interface optimized for non-visual use. In Proceedings of the 27th international conference on Human factors in computing systems, CHI ’09, pages 1829–1838, New York, NY, USA, 2009. ACM.

[11] E. Bursztein, R. Bauxis, H. Paskov, D. Perito, C. Fabry, and J. C. Mitchell. The failure of noise-based non-continuous audio captchas. In Proceedings of 2011 IEEE Symposium on Security and Privacy (Oakland’11), 2011.

[12] E. Bursztein, R. Beauxis, H. S. Paskov, D. Perito, C. Fabry, and J. C. Mitchell. The failure of noise-based non-continuous audio captchas. In Proceedings of the 2011 IEEE Symposium on Security and Privacy. IEEE Computer Society, 2011.

[13] E. Bursztein, S. Bethard, C. Fabry, D. Jurafsky, and J. C. Mitchell. How good are humans at solving captchas? a large scale evaluation. In Proceedings of 2010 IEEE Symposium on Security and Privacy (Oakland’10), 2010.

[14] T.-Y. Chan. Using a text-to-speech synthesizer to generate a reverse turing test. Tools with Artificial Intelligence, IEEE International Conference on, 0:226, 2003.

[15] K. Chellapilla, K. Larson, P. Simard, and M. Czerwinski. Designing human friendly human interaction proofs (hips). In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI ’05,

pages 711–720, New York, NY, USA, 2005. ACM.

[16] K. Chellapilla and P. Simard. Using machine learning to break visual human interaction proofs (hips). In In Advances in Neural Information Processing Systems, pages 265–272, 2005.

[17] M. Chew and J. D. Tygar. Image recognition captchas. In In Proceedings of the 7th International Information Security Conference (ISC), pages 268–279, 2004.

[18] R. Datta, J. Li, and J. Z. Wang. Imagination: a robust image-based captcha generation system. In Proceedings of the 13th annual ACM international conference on Multimedia, MULTIMEDIA '05, pages 331–334, New York, NY, USA, 2005. ACM.

[19] A. S. El Ahmad, J. Yan, and L. Marshall. The robustness of a new captcha. In Proceedings of the Third European Workshop on System Security, EUROSEC '10, pages 36–41, New York, NY, USA, 2010. ACM.

[20] J. Elson, J. R. Doucerur, J. Howell, and J. Saul. Asirra: A captcha that exploits interest-aligned manual image categorization. In Proceedings of the 14th ACM conference on Computer and communications security, CCS '07, pages 366–374, New York, NY, USA, 2007. ACM.

[21] H. Gao, H. Liu, D. Yao, X. Liu, and U. Aickelin. An audio captcha to distinguish humans from computers. In Proceedings of the 2010 Third International Symposium on Electronic Commerce and Security, ISECS '10, pages 265–269, Washington, DC, USA, 2010. IEEE Computer Society.

[22] P. Golle. Machine learning attacks against the asirra captcha. In Proceedings of the 15th ACM conference on Computer and communications security, CCS '08, pages 535–542, New York, NY, USA, 2008. ACM.

[23] R. Gossweiler, M. Kamvar, and S. Baluja. What's up captcha?: a captcha based on image orientation. In Proceedings of the 18th international conference on World wide web, WWW '09, pages 841–850, New York, NY, USA, 2009. ACM.

[24] R. Guha, R. McCool, and E. Miller. Semantic search. In Proceedings of the 12th international conference on World Wide Web, WWW '03, pages 700–709, New York, NY, USA, 2003. ACM.

[25] P. Matthews and C. C. Zou. Scene tagging: image-based captcha using image composition and object relationships. In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, ASIACCS '10, pages 345–350, New York, NY, USA, 2010. ACM.

[26] G. Mori and J. Malik. Recognizing objects in adversarial clutter—breaking a visual captcha. In In Proceedings of the Conference on Computer Vision and Pattern Recognition, 2003.

[27] Y. Rui and Z. Liu. Excuse but are you human? In Proceedings of the eleventh ACM international conference on Multimedia, MULTIMEDIA '03, pages 462–463, New York, NY, USA, 2003. ACM.

[28] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford. Captcha: Using hard ai problems for security. In In In Proceedings of Eurocrypt, Vol. 2656, pages 294–311, 2003.

[29] L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart automatically. Commun. ACM, 47:56–60, February 2004.

[30] J. Yan and A. S. El Ahmad. A low-cost attack on a microsoft captcha. In Proceedings of the 15th ACM conference on Computer and communications security, CCS '08, pages 543–554, New York, NY, USA, 2008. ACM.

[31] J. Yan and A. S. El Ahmad. Usability of captchas or usability issues in captcha design. In Proceedings of the 4th symposium on Usable privacy and security, SOUPS '08, pages 44–52, New York, NY, USA, 2008. ACM.

[32] B. B. Zhu, J. Yan, Q. Li, C. Yang, J. Liu, N. Xu, M. Yi, and K. Cai. Attacks and design of image recognition captchas. In Proceedings of the 17th ACM conference on Computer and communications security, CCS '10, pages 187–200, New York, NY, USA, 2010. ACM.