

A Proposal For Improvements of Image Based CAPTCHA

A Proposal for Improvements of Image Based CAPTCHA

Dennis Egen

Rutgers University

December 23, 2009

degen@camden.rutgers.edu

1. Summary

CAPTCHAs are commonly used security measures on the internet that prevent automated programs from abusing online services. They do so by asking humans to perform a task that computers cannot yet perform, such as recognizing shapes, deciphering distorted characters, discerning objects or animals in images or solving puzzles. CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart. As we will see, many attacks have been attempted to break or beat CAPTCHAs with some success. We will look at these unique, creative and even resourceful attempts and subsequent countermeasures and propose a new system of our own, building on some recent successes.

2. Background

The idea of CAPTCHAs or reverse Turing tests, as they are sometimes called, was first proposed in 1996 by Moni Naor. Naor stated that what should be used is “those tasks where humans excel in performing, but machines have a hard-time competing with the performance of a three year old child.” Among those tasks proposed were: Gender recognition (we will explore this more in the proposal), Facial Expression Understanding, Handwriting understanding, and Filling in words (Naor 1996).

In 1997, shortly after Naor’s paper was published, Yahoo! was having huge problems with spammers using bots to signing up for free email accounts. Yahoo! went to Carnegie Mellon and asked them for help. By 2000, the Carnegie Mellon team had invented the first CAPTCHA. The CAPTCHA caught on, and now it’s all over the Web. Luis von Ahn, an assistant professor at Carnegie Mellon who was part of the original CAPTCHA team, estimates that people fill out close to 200 million CAPTCHAs a day. (Grossman 2008)

By 2008, most major forms of CAPTCHA had been attacked with a success rate ranging from 30% to 60% (Prasad) (Yan and Ahmad, A Low-cost Attack on a Microsoft CAPTCHA 2008)

CAPTCHAs for the Common Good

In September 2008 Von Ahn and his team at Carnegie Mellon released their latest twist on CAPTCHA in a paper published in the journal Science. The goal of reCAPTCHA was to utilize all of the time humans spend filling out CAPTCHAs and channel it for a useful purpose.

There are many ongoing projects to digitize books that were published before computers or the internet. These efforts will allow users to search these texts online on sites like

A Proposal For Improvements of Image Based CAPTCHA

Google Books. Also, these efforts will preserve human knowledge and make these books more accessible. This is an arduous process that involves OCR (Optical Character Recognition) software with the help of human transcribers. According to the research OCR software can fail to recognize up to 20% of words in some of the older more faded texts. Humans, on the other hand can be 99% accurate using what is called the “key and verify” method (This involves two humans, typing the text independently and then identify and resolve discrepancies) (Luis von Ahn 2008).

reCAPTCHA aims to greatly improve this process by using these hard-to-recognize texts as CAPTCHA images. The user is presented with the image to recognize along with a control image. This two-image system is used to weed out extraneous texts that are unreadable, even by humans. If an incorrect answer (along with a correct control answer) is given a certain amount of times for a given image, it is taken out of circulation. Accuracy of reCAPTCHA system was in the 99% range, the same as the industry standard key and verify. This system is the most secure text based CAPTCHA system since it ensures that the words it uses are in fact unrecognizable by OCR software. As we will see below, Microsoft’s ASIRRA project also uses CAPTCHAs for humanitarian purposes.

Usability and Security Issues

Anyone who has had to strain their eyes or squinted trying to fill one of these out can tell you: CAPTCHAs are not user friendly. In fact, studies have shown that CAPTCHAs can be a barrier to entry and negatively affect usability. CAPTCHAs can be particularly difficult for non-native speakers (Yan and Ahmad, Usability of CAPTCHAs 2008). Consequently, studies show this can directly affect conversion rates. A particular study conducted over the course of three months involving fifty websites showed that CAPTCHA’s inherent barrier to usability was responsible for a loss of 3.2% of conversions (in this case sign-ups). This could be very detrimental to revenue for a firm. (chenry 2009)

One could argue that reCAPTCHA is even less usable because the quality of the images is less controlled. Some of the texts just cannot be read by a human.

Image Recognition CAPTCHA

In order to tackle the inherent usability issues that come with text based CAPTCHAs some work has been done to investigate image recognition CAPTCHAs. The first of these was probably KittenAuth . KittenAuth requires the user to select a kitten (or other stated type of animal) from an array of thumbnail images of assorted animals. The images (and the challenge questions) can be customized, for example to present questions and images which would be easily answered by a forum's target user-base. The user is presented with a grid of 9 pictures to choose 2 kitten images from. This ensures a brute force attack (Just guessing 2 images over and over again) has a very small success rate. In

A Proposal For Improvements of Image Based CAPTCHA

this case the average success rate would be 1 out of (9 choose 2) times or $1/36$ (.0277). KittenAuth has limitations: KittenAuth, by default, only comes with a library of 42 images. Also, by design, those images do not change. This means, an attacker could write a script to download all of the images (even if the library had a few thousand images), compute a hash value for each, and mark each one (by hand) as a kitten or non-kitten. They could then use this to easily automate downloading each image, comparing hash values and identifying its *kittleness*.

In order to solve the problem of a small image library Microsoft Research started the ASIRRA project. According to their site "(<http://research.microsoft.com/en-us/um/redmond/projects/asirra/>): "ASIRRA is a human interactive proof that asks users to identify photos of cats and dogs. It's powered by over three million photos from our unique partnership with Petfinder.com". Users are given the ability to click a link to adopt a pictured cat or dog at petfinder.com. ASIRRA may be the most promising image recognition CAPTCHA (Human Interaction Proof as Microsoft calls it). However, it is not without its limitations and vulnerabilities. Recently a group of researchers used machine learning techniques to identify different patterns in images of dogs versus cats. This turned out to be effective 10.3% of the time (Golle 2008). While this is not really a feasible technique for even advanced spammers it does present vulnerability. Another interesting attempt at overcoming the limited image library problem was made by HotCAPTCHA, which used an API provided by the [awful] site www.hotornot.com. This used a much more subjective approach of asking the user to pick the three attractive (hot) people from a grid of hot and not people. The choices would be matched with the ratings of the user-base of hotornot.com. Hotornot.com has since discontinued their API. Also, as it turns out women rating men is way more subjective than the reverse (Go figure) (Wood and Brumbaugh Jun 2009). More troublesome, the implementation was flawed in that the entire call to the API was done client side so an attacker could easily script against each image by retrieving the rating.

Human Readers

There are several companies that advertise CAPTCHA solving on the internet. Basically, CAPTCHAs are sent to Human Readers who break the CAPTCHA manually. Sometimes they are rewarded monetarily. Sometimes the reward is more creative: There have been reports of pornography sites presenting customers with CAPTCHAs that they must break to see the images they wish to see. Thus, the CAPTCHA is broken and access can be granted to the victim site. Human readers should not be considered a security concern for CAPTCHAs since they are providing the expected results. That is, humans are confirmed as humans. Other defenses against these attacks should be considered but will not be discussed here.

Accessibility

A Proposal For Improvements of Image Based CAPTCHA

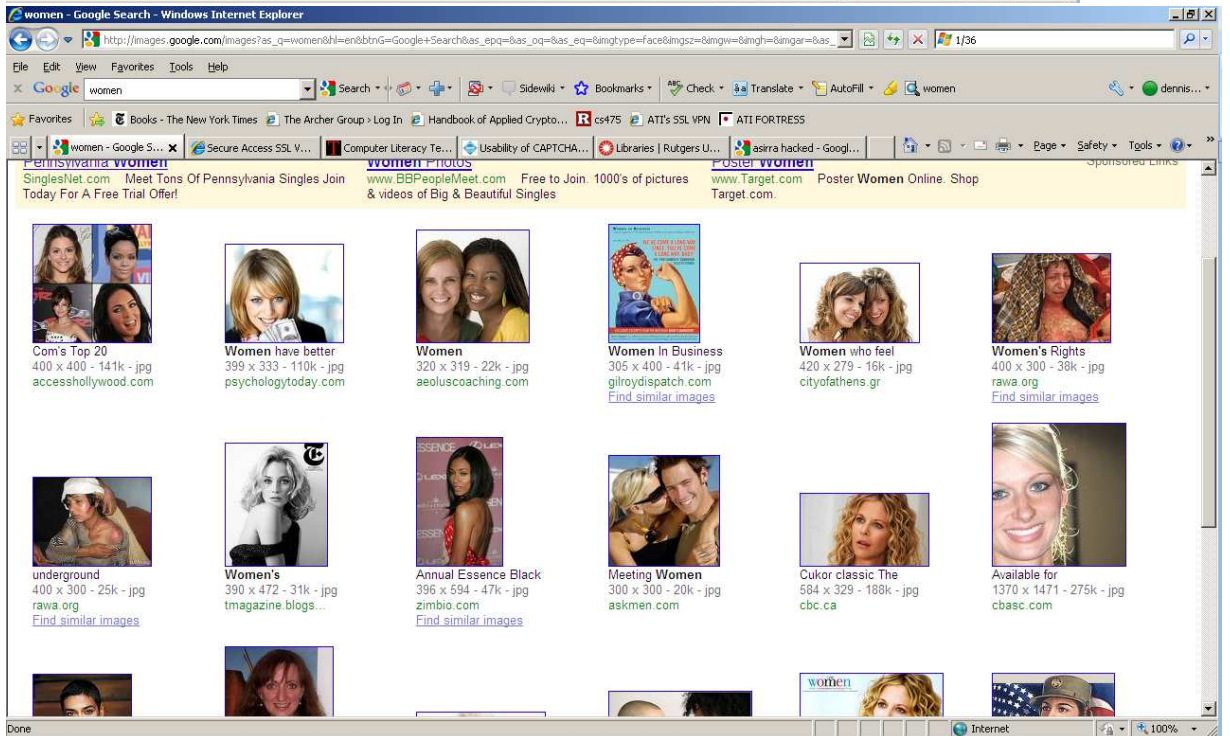
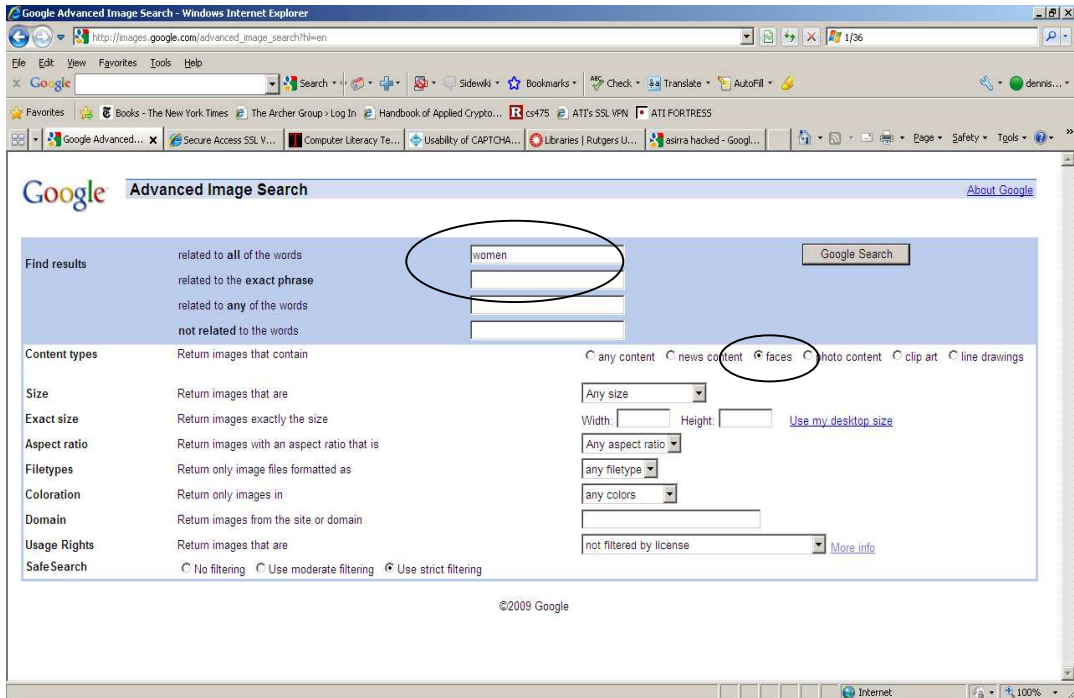
Since all CAPTCHAs rely on visual perception, they are not usable by those with impaired vision. Thus far, the only resolution to this that keeps with the true definition of a CAPTCHA is audio CAPTCHA. Basically, a user can opt to solve a voice-recognition CAPTCHA that presents the user with an audio clip of numbers, letters or words. So far image recognition CAPTCHAs like KittenAuth are not accessible to users with visual disabilities. This could keep these CAPTCHAs from being used widely as many large commercial sites insist on an acceptable level of accessibility.

3. Proposal

In order to build on current progress and remove limitations in image recognition CAPTCHAs we propose a system which greatly increases the image library size while decreasing the likelihood of machine learning type attacks. Also, we must maintain the ease of use inherent in the ASIRRA interface. Quite probably, it would be more difficult to produce the kind of results that were achieved in the machine learning study if the task was to distinguish male faces from female faces (rather than cats from dogs). This is not a proven fact but a conjecture based on the relative similarities of human faces (when compared to those of cats and dogs). Obviously, this would need to be tested, but this assumption is no less tested than the assumptions made by KittenAuth or ASIRRA early on.

Searching Google Images for “women” or “men” (using the advanced search feature to specify “faces” as criteria) returns millions of results that could be made into a very reliable library for our CAPTCHAs.

A Proposal For Improvements of Image Based CAPTCHA



The Google Search API provides an interface that could be used to gather these images (programmatically) for server side processing. As a simple proof of concept, we downloaded a windows executable that downloads thousands of images using the Google

Search API. The images should be logically 'stored' in a database to keep track of quality and use of images. The initial image load could take place over a few weeks or so. Then, the library could be updated on an ongoing basis.

Possible risks and limitations

Integrity of the images:

As can be seen from the search results, the image results are not completely reliable. That is, some images are men, some have men and women and some aren't even human. However, an unscientific review of about 100 images shows that about 80-90% are easily recognizable as one gender or the other. This could be mitigated by a reCAPTCHA type system where users are presented with an additional wild card image along with the real CAPTCHA to evaluate. Suspect images could be presented to the user and their response to it correlated with their responses to the 'known good' images. Consequently, these poor images could be ranked lower and lower until they are no longer used.

Reverse engineering of images:

Although less likely than the attack we saw against KittenAuth, since the images are public, a dictionary attack is possible. To avoid use of hash values against the system the images should be updated (imperceptibly) from time to time to change the hash value.

Attacks against the rating system:

Malicious scripts could be written to increase the rank of bad images, causing the system to malfunction. This was tried unsuccessfully against reCAPTCHA in early 2009:

<http://musicmachinery.com/2009/04/27/moot-wins-time-inc-loses/>

Legality

We have not investigated the legality of using these images in this way.

4. Conclusion

We have now explored the many incremental improvements made to CAPTCHAs over time. Herein, we propose another CAPTCHA system that builds upon known good techniques used in reCAPTCHA, ASIRRA, and KittenAuth. Our system attempts to solve the problems of usability experienced in text based CAPTCHAs, especially reCAPTCHA. We attempt to reduce exposure to machine learning attacks that we say against ASIRRA. We do this by keeping our choices visually similar. We greatly increase image library size by tapping into Google's vast library of images.

Bibliography

chenry. *SEOMoz.org*. July 17, 2009. <http://www.seomoz.org/blog/captchas-affect-on-conversion-rates>.

A Proposal For Improvements of Image Based CAPTCHA

Golle, Philippe. "Machine learning attacks against the Asirra CAPTCHA." *Conference on Computer and Communications Security*. Alexandria, VA, USA: ACM, 2008. 535-542.

Grossman, Lev. "Computer Literacy Tests: Are You Human?" *Time*, June 5, 2008.

Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, Manuel Blum. "reCAPTCHA: Human-Based Character." *Science*, 2008: 1465-1468.

Naor, Moni. "Verification of a human in the loop or Identification via the Turing Test*." 1996.

Prasad, Sumeet. *WebSense*. <http://securitylabs.websense.com/content/Blogs/2919.aspx>.

Wood, Dustin, and Claudia Chloe Brumbaugh. "Using revealed mate preferences to evaluate market force and differential preference explanations for mate selection." *Journal of Personality and Social Psychology*, Jun 2009: 1226-1244.

Yan, Jeff, and Ahmad Salah El Ahmad. "A Low-cost Attack on a Microsoft CAPTCHA." 2008.

—. "Usability of CAPTCHAs." *ACM International Conference Proceeding Series*. Pittsburgh, PA: ACM, 2008. 44-52.