CAPTCHAs with a Purpose

Abstract

In this paper, we develop a new genre of Captchas called CAPTCHAs with a purpose. These CAPTCHAs apart from having its applications serve some useful purpose. reCAPTCHA is one such Captcha developed at Carnegie Mellon University. It helps to digitize books. Another such Captcha is Asirra developed at Microsoft which provides homes for homeless animals. In this paper, we present Time based, Sentence based, Human Emotion based CAPTCHAs which have range of other useful purpose such as measuring reaction time of people, promoting news, general knowledge facts, jokes among people while engaging in routine activities such as checking email. Also they can be used for conducting online polls on a very large scale.We also present a new scheme which renders attack on CAPTCHAs useless and make old CAPTCHAs reusable and help in using CAPTCHAs which might serve some practical purpose which otherwise might be vulnerable to use. This system also enables to use different 'CAPTCHAs with a purpose' in conjunction with each other. At present most websites deploy only a single algorithm reCAPTCHA whose practical purpose is to digitize books, thus have a very limited practical utility and benefit to the human community. This system can thus, broaden the application domain of CAPTCHAs.

1. Introduction

A CAPTCHA or an HIP is a program that protects websites against bots by generating and grading tests that humans can pass but current computer programs cannot[1]. For example, humans can read distorted text, but current computer programs can't. The term CAPTCHA (for Completely Automated Public Turing Test To Tell Computers and Humans Apart) was coined in 2000 by Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford of Carnegie Mellon University.

1.1 Applications : CAPTCHAs are used to prevent automated software from performing actions which degrade the quality of service of a given system, whether due to abuse or resource expenditure. CAPTCHAs can be deployed to protect systems vulnerable to e-mail spam, such as the webmail services of Gmail, Hotmail, and Yahoo! Mail.CAPTCHAs found active use in stopping automated posting to blogs, forums and wikis, whether as a result of commercial promotion, or harassment and vandalism. CAPTCHAs also serve an important function in rate limiting, as automated usage of a service might be desirable until such usage is done in excess, and to the detriment of human users. In such a case, a CAPTCHA can enforce automated usage policies as set by the administrator when certain usage metrics exceed a given threshold. The article rating systems used by many news web sites are another example of an online facility vulnerable to manipulation by automated software.

1.2 CAPTCHAs with a Purpose : In this paper we develop a new genre of Captchas called CAPTCHAs with a purpose. These Captchas apart from having its applications serve some useful purpose. reCAPTCHA [2] is one such Captcha developed at Carnegie Mellon University. It helps to digitize books. Another such Captcha is Asirra [3] developed at Microsoft. In this paper, we present Captchas which have range of other useful purpose such as measuring reaction time of people, promoting news, general knowledge facts, jokes among people while engaging in routine activities such as checking email, conducting online polls on a very large scale.

2.Time Based Captcha

Main idea behind this CAPTCHA is to make use of challenges which humans can solve quickly but computers take some time to solve. These CAPTCHAs expire, if they don't receive a response in a particular time interval.

1)Splash random alphabets on random positions on a screen at different time instants. Alphabets will appear for an instance and then disappear. User will have to type alphabets in a sequence they appeared.

2) Display simple images in a sequence. Images will appear for an instance and then disappear. User will have to describe images in a sequence they appeared.Sample images are such as those labelled by an ESP game [4].

PURPOSE : As user will have to quickly respond to the alphabets displayed as they appear and disappear, this can be used to check whether user is able to react to the alphabets presented in a certain time frame, so it can be a suitable means to measure certain reaction times of human beings. A good data about human beings which is otherwise difficult to obtain and is very useful.

Analysis: There is some challenge in implementation of these CAPTCHAs, if implementation is done using flash, then flash file might be downloaded.

A news headline presented to the user

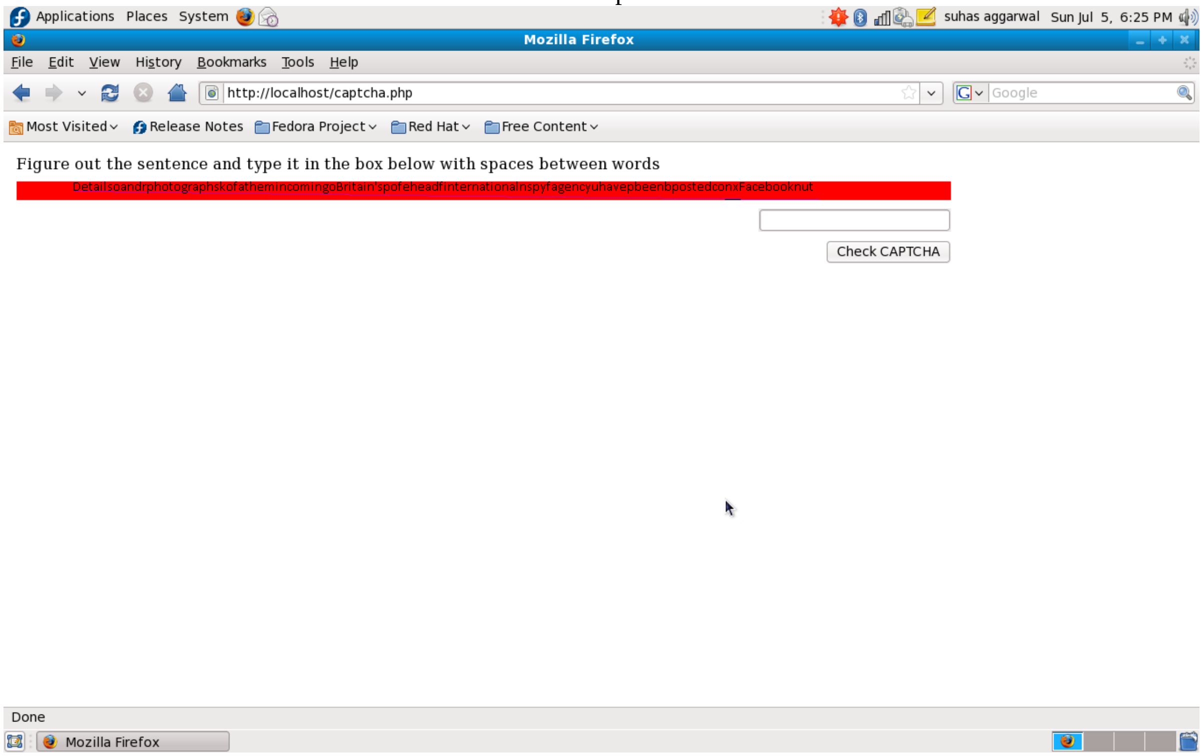

Applications Places System
suhas aggarwal Sun Jul 5, 6:25 PM
Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost/captcha.php
Google
Most Visited
Release Notes
Fedora Project
Red Hat
Free Content
Check CAPTCHA
Done
Mozilla Firefox

Figure out the sentence and type it in the box below with spaces between words

DetailsoandrphotographskofathemincomingoBritain'spofeheadfinternationalnspyfagencyuhavepbeenbpostedconxFacebooknut

Figure 1: Prototype of sentence based captcha

3.Sentence Based Captcha

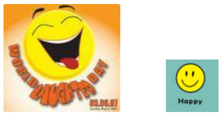
A sentence is selected. Two random words are selected from the sentence and are swapped. A Random alphabet is filled in each whitespace present in the sentence. Sentence is degraded by crippling its text (GIMPY) for interfaces with GUI support or by using Text - Graphics character based CAPTCHAs [5] for interfaces with text support only such as SSH accounts. Challenge is to guess the sentence and write the sentence in correct format by inserting whitespaces between words. A possible reversible attack can be to write a program which can guess correct words in the sentence and perform a search over the web and identify the correct sentence from the results obtained. A new degrading scheme which can be useful in this scenario is to omit out certain alphabets from the words and present a sentence composed of partially 'eaten up' words.

PURPOSE : Educating people. Sentence can be general knowledge facts, thoughts of the day, recent headlines. (Imagine how cool! will it be, if one can catch up with the latest news while just checking regular emails!)

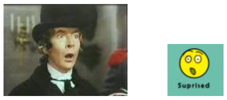
Analysis : There is an argument that Sentence based CAPTCHA is bit difficult for humans.It might also be possible to attack it using natural language processing techniques using language model techniques for example. Sentences making use of common sense facts might be more useful to follow taking into account that machines don't have access to this information but then this CAPTCHA will loose its practical utility. How to use this CAPTCHA so that it can serve its purpose is discussed later in the paper.

4.Human Emotion Based Captcha

A statement or a graphic is displayed to the user arousing human emotion. User has to choose an emoticon describing his feeling as an answer .



(a)



(b)

(c)

(d)

(e)

1)Not all scars show, not all wounds heal

Sometimes you can't always see

The pain someone feel Sympathetic

2)Are children who act in rated 'R' movies allowed to see them? Happy

3)"The internet is a great way to get on the net."

- Bob Dole, Republican presidential candidate Thoughtful

4)The man who smiles when things go wrong has thought of someone to blame it on. Sheepish

5)Hewlett Packard's first product was an automatic urinal flusher Interested

6)In the late '90s, Microsoft secretly developed its own version of Linux, but shelved it after quality control researchers deemed it "too stable". Curious

PURPOSE : Can be indirectly used for conducting online polls on a very large scale. When an important poll is being conducted sometimes, all answers given by the users will be marked as acceptable and their different reactions corresponding to an event can be obtained. It won't affect the security of CAPTCHA as poll is conducted sometimes and it is not possible to know when it will be conducted.

Example -

Austrian-born actor Arnold Schwarzenegger has won the race to become governor of California, pledging to become a "governor of the people."

Poll –

Confident - 35%

Satisfied - 35%

Disapproving - 30%

Analysis : .Security analysis of this CAPTCHA is similar to that of Asirra CAPTCHA. It is argued that this CAPTCHA is based on the security of the database but as are there are plenty of such images available from different sources on the web, it is an easy task to upgrade the database frequently, hence it becomes difficult to attack this CAPTCHA. KittenAuth suffered from a database attack because its database made use of only 42 images but there can be millions of images which can be used for this CAPTCHA. Moreover, the technique used is not merely image recognition as employed in CAPTCHAs such as Asirra which have been attacked, one has to depict emotion displayed by images.

5.Scenario Based Captcha

Most of the Captchas employed today are based on the difficulty of image recognition, such as identifying degraded words etc. With the advancement in field of computer vision these Captchas have already been broken. Idea behind this CAPTCHA is to utilize the analytic and understanding capability of humans rather than merely recognizing objects. A scenario is presented to the user. Here are few figures depicting a scenario.

(a)

(b)

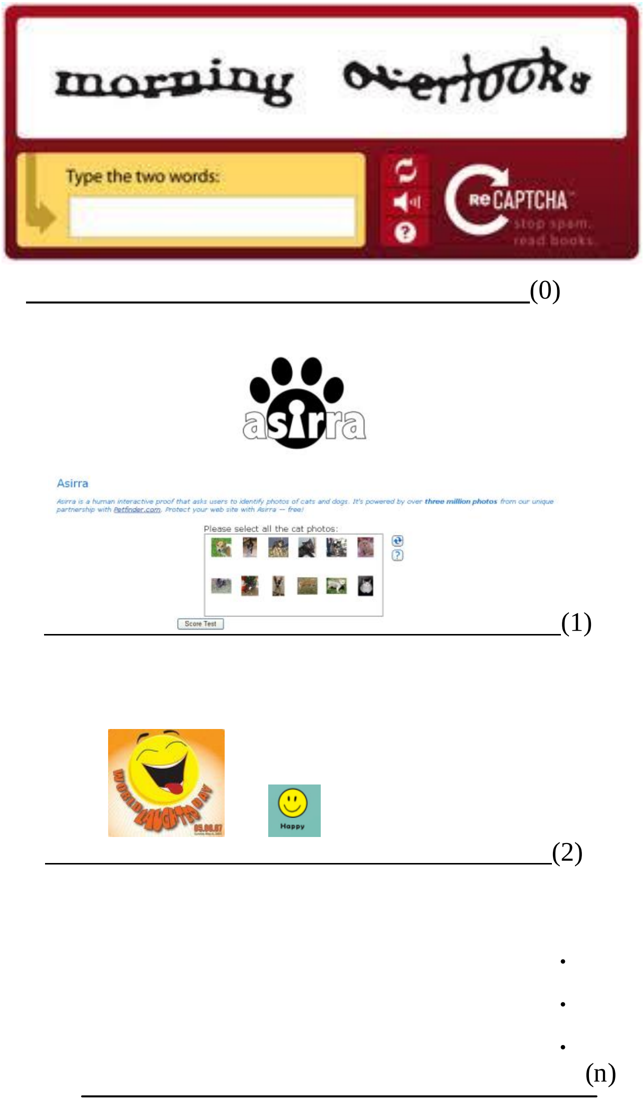
(c)

(d)

These figure illustrates various scenarios such a monkey eating a banana, a couple dancing, a student studying at night, two people wrestling. A graphic like this is presented to the user. He has to type in the expression depicting a scenario. Answer is a simple expression such as those described above. There can be multiple answers possible as well such as a chimp eating, a couple performing a ballroom dance, a student studying for his exam etc. These expressions are simply paraphrases and can be easily tackled due to advancements in computational linguistics today. Paraphrase recognition systems such as iSTART [6] can be used for recognizing paraphrases. This Captcha uses a step higher than object recognition that is identifying the objects present in the picture as well as how they are related. By using object recognition techniques we can analyse various objects present in the images but identifying relationship among them is a hard AI problem to crack. Moreover they may be related in multiple ways and unique relationship can only be identified from image shown. Distortions such as blurring images can also be added to make it complex for computers to tackle.

6.Security of CAPTCHAs, Means of Reusing Old CAPTCHAs and Enabling them to serve their purpose.

Most of the websites today use a single CAPTCHA to secure the login interface. As there are plenty of CAPTCHAs available today, different CAPTCHA problems can be used to secure a single login interface. Set of different CAPTCHA problems is maintained and a random CAPTCHA is selected from these CAPTCHAs before a login prompt is granted.



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. Protect your web site with Asirra — free!

Please select all the cat photos:

Figure 2 : Selection from Different CAPTCHAs

Above figure, illustrate different CAPTCHA problems generated from different CAPTCHA scripts. A Random

CAPTCHA script, (A random CAPTCHA problem) is selected each time before giving login prompt to the user. If a login is unsuccessful, a random CAPTCHA problem is selected and presented to the user. As CAPTCHAs of different genre are presented randomly, this makes it very difficult to attack these CAPTCHAs. Even if a program is written to attack one CAPTCHA problem, devising an online password guessing attack becomes infeasible as after one unsuccessful try, a different random CAPTCHA problem will be presented to the user and attacking program should be capable to solve that as well. This problem, introduces a new problem of identifying CAPTCHAs, before attacking a CAPTCHA, attacking program will have to identify which CAPTCHA, it is as well.

How it aids in enabling CAPTCHAs serve some practical purpose?

CAPTCHAs such as sentence based CAPTCHA presented earlier might be bit annoying to people if presented to people everytime, but as in the above technique, a random CAPTCHA is presented each time, people might encounter it sometimes and might be willing to solve it to gain some useful knowledge. Moreover, as random CAPTCHAs are presented to the user, it is not possible to attack them as a graphic can be a sentence, image with different challenges such as figuring out sentence, typing emotion corresponding to image, typing letters, so each CAPTCHA has a different solution which can be figured out by humans more easily but is difficult for computer as it does not know what to give as a answer.

Possible attack –

HTML protectors must be deployed to prevent automated CAPTCHA classification by content analysis of html, javascript returned from servers. Some softwares such as by Antssoft, creabit are available for this purpose.

7. Increasing the Usability of CAPTCHAs using Partial Credit and Token Bucket Algorithm

Partial credit scheme is discussed in Asirra[3]. It can also be used to make CAPTCHAs bit easy for people which are otherwise difficult. For eg- If in a Sentence based CAPTCHA, sentence being presented to the user is bit difficult and he is able to figure out say 5 out 7 words , he is given partial credit and a second CAPTCHA is presented to the user, this scheme can be combined with the above the system. Second CAPTCHA presented to the user may or may not be Sentence Based CAPTCHA, for example it might be Text based CAPTCHA and if person is able to get that partially correct as well, he is believed to have solved CAPTCHA challenge completely. Also, Token bucket schemes [3] which makes use of fact that bots used small number of IP addresses that submit a very large number of incorrect responses, interspersed with a much smaller number of correct responses can be used to increase the security of CAPTCHAs.

8.Conclusion

In this paper, we have developed and discussed the need to use 'CAPTCHAs with a purpose' which serve some other useful, practical purpose so that human hours spend in solving CAPTCHAs are channelised for benefit of human community. We have also shown a system which enables CAPTCHAs having different 'practical purpose' to be used in conjunction with each other. At present, most websites deploy reCAPTCHA which just serves one 'practical purpose' to help digitize books, this CAPTCHA can be used in conjunction with other 'CAPTCHAs with a purpose' to broaden the application domain of CAPTCHAs and make CAPTCHA solving more interesting and beneficial to the human community.

10.References

[1]Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford. CAPTCHA: Using Hard AI Problems for Security. Advances in Cryptology, *Eurocrypt 2003*. pp 294-311

[2]Luis von Ahn, Ben Maurer, Colin McMillen, David Abraham and Manuel Blum. reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, September 12, 2008. pp 1465-1468

[3]Asirra: A CAPTCHA that Exploits Interest-Aligned Manual Image Categorization from *Microsoft Research*

[4]Luis von Ahn. Games With A Purpose. *IEEE Computer Magazine*, June 2006. pp 96-98.

[5]Dailey, M. and Namprempre, C. 2004. A text graphics character CAPTCHA for password authentication. In Proceedings of the *IEEE Region 10 Conference (TENCON, Nov. 21--24)*, 45--48.

[6]Boonthum, C. (2004). iSTART: Paraphrase Recognition. *Proceedings of the Student Research Workshop: ACL 2004. 42nd Annual Meeting of the Association of Computational Linguistics,* Barcelona, Spain. pp 31-36.

[7]Luis von Ahn, Shiry Ginosar, Mihir Kedia and Manuel Blum.Improving Accessibility of the Web with a Computer Game. *ACM Conference on Human Factors in Computing Systems*, CHI Notes 2006. pp 79-82.

[8]Luis von Ahn, Mihir Kedia and Manuel Blum. Verbosity: A Game for Collecting Common-Sense Knowledge. *ACM Conference on Human Factors in Computing Systems,* CHI Notes 2006. pp 75-78.

[9]Luis von Ahn, Ruoran Liu and Manuel Blum. Peekaboom: A Game for Locating Objects in Images. *ACM Conference on Human Factors in Computing Systems*, CHI 2006. pp 55-64.

[10]Baird, H. S. and Riopka, T. 2005. Scatter Type: A reading CAPTCHA resistant to segmentation attack. In Proceedings of the SPIE/IS&T Conference on Document Recognition and Retrieval XII (San Jose, CA, Jan.).

[11]"Collaborative filtering CAPTCHAs." M. Chew and J. D. Tygar. In *Human Interactive Proofs: Second International Workshop (HIP 2005),* eds. H. Baird and D. Lopresti, Springer, May 2005, pp. 66-81.

[12]Phish and HIPs: Human interactive proofs to detect phishing attacks." R. Dhamija and J. D. Tygar. In *Human Interactive Proofs: Second International Workshop (HIP 2005),* eds. H. Baird and D. Lopresti, Springer, May 2005, pp. 127-141.

[13] Deapesh Misra , Kris Gaj, Face Recognition CAPTCHAs, Proceedings of the *Advanced Int'l Conference on Telecommunications and Int'l Conference on Internet and Web Applications and Services*, p.122, February 19-25, 2006