A Study on User Friendly Approach : CAPTCHA

Vanya Saxena(M.Tech)
Kanpur Institute of Technology, Kanpur
er.vanyasaxena@gmail.com

Abstract : Captcha stands for completely automated public turing test to tell computer and human apart. Due to the increasingly growth of computers and internet, the security of the web applications plays a vital role. Many web applications facing a threat of internet bots. In most of the time, we have to register in order to use the facilities of the internet. Registration is basically identification of the receiving person. But today, hackers are performing the automatic registration on the internet sites by using computer program; this leads the wastage of resources of the sites. This can be easily handled by captcha. Captcha is the standard for securing web applications from internet bots and almost all the registration web forms use this test. During the last decade many researchers has done a lot of work on captcha systems. This paper is a collective analysis of work done on captcha system. In the first section a typical introduction about the captcha has been discussed and in the section 2 discussed about the applications of captcha. The section 3 discussed about the characteristics of captcha. In the last in section 4 discusses about the types of captcha and how they are used according to the need.

I. Introduction

Captcha. A(completely automated public turing test to tell computers and humans apart). It is a test to differentiate a human users from the computer system. Captchas are also known as reverse turing test. The Captchas have been used since the late 1990s to distinguish between human and machines(bots) in the online context. Captcha are one such mechanism act as a first line of defense at the interface between human users and the internet. A good captcha easy for humans to solve but next to impossible for machines to solve. Today, all most all the internet users have login accounts for internet sites and these sites require only the registration by human users but unfortunately some automated computer programs to enter these sites and use their resources through false registration. A good captcha must not only be user friendly but also robust enough to resist computer programs that attackers/hackers write to automatically pass captcha test. Captcha resist the automatic registration over the internet and provides the smooth registration process. It is basically used for a security purposes in free e-mail services, online polls, search engine bots, money transactions etc. This test is required to ensure that in online polls systems the vote is submitted by the humans not by the computer programs. There is a check that money transaction is done by the humans only. Captcha is a challenge response test used by computer programs to guarantee that the response is not generated by another computer (or program) [1].

The term Captcha was introduced by John Langford of Carnegie Mellon University[2] but the ground work was already done by Mani Noar who introduced the concept of turing test to identify the difference between a human and bot in 1996[3]. The CMU team is one of the mostly active research team in captcha research from that point onwards. They have developed a number of concrete Captcha system[4] such as *Gimpy*, *Bongo*[5], *Pix*, *Animal Pix* and *recaptch*a[6]. Captcha system are very useful to prevent brute force attacks and Daniel of services. Captcha test is also useful to avoid unwanted comments on the blogs and also restricting advertising spam by intended computer programs.

II. Applications of Captcha

Captchas are important component of web application's security. Captchas have a variety of applications. we are discussing them here.

1. ***Free E-mail services*** : Several companies(such as Yahoo!, Google, Rediff, Microsoft) offer free e-mail services, most of which suffer from a specific type of attack : bots that sign up for thousand of e-mail accounts every minute[7]. This situation can be avoided or improved by requiring users to prove that they are human before they can get a free email account. Rediff, for instances, uses a captcha to prevent bots from downloading a files from email account and Yahoo! registering for email accounts.

2. ***Search Engine bots*** : Owners of some websites do not want their sites to be indexed by search engines. They uses an HTML tags to prevent search engine bots from reading web pages. However the tags does not guarantee that the bots will not read the pages and these tags will not stop bots from indexing those web pages. In order to truly guarantee that bots will not enter a web site, Captchas are needed.

3. ***Worms and Spam*** : Captcha offers a preventive solution against email worms and spam: email clients can adapt to a policy such as it only accepts an email if it knows there is a human behind the other computer from where that particular email is sent.

4. ***Online Polls*** : Captcha also offers a preventive solution for online voting systems. As is the case with most online polls, IP addresses of voters were recorded in order to prevent single users from voting more than once.

5. ***Online Games*** : Another application of captcha is online games[8][9][10] where it is preventing web robots from playing games.

6. ***Preventing Dictionary Attacks*** : The idea is to prevent a computer from being able to iterate through the entire space of passwords by requiring a human to type the passwords. Pinkas and Sander[11] have suggested using Captchas to prevent dictionary attacks in password based systems.

III. Characteristics of Captcha

There are several types of captcha used in the recent past like Text Captchas, Image Captcha and Sound Captcha. The peoples used any type of captcha methods according to their application's requirement. The characteristics[7] of these captcha's are:

1. ***Easy for human*** : This test is taken by a human user and this test should be quickly (the response time is typically within 30 seconds).

2. ***Automation and grade-ability*** : This test is generated and graded by a machine and the machine will not be able to solve the test. This is the main requirement of a Captcha.

3. ***Robustness when database publicized*** : When database is publicized, the test should be difficult to attack from which the test is generated.

4. ***Challenging and hard for Machine*** : As the captchas is based on the hard AI system. The test should be based on a well -known hard AI problem and the best existing techniques should be far from solving the problem.

5. ***Universality*** : The test should be independent of user's language, educational background and geographical areas. This is motivated by practical considerations, and especially important for companies like Google, Microsoft and Yahoo! those having international customers. Strictly saying that, no captcha test can be absolutely universal, as there are no two humans who are same in this world. So, we can make assumptions. For instance, it is reasonable to assume that a human who is using a computer knows the 26 English alphabets and the 10 digits.

IV. Types of Captchas

I. Text based Captcha

Fig 1: Text based Captcha

The most widely used captcha is the text based captcha used in web applications. In this captcha system the server renders set of characters after distorting the text. Many websites uses their own captchas. The another type of interesting captcha is reCaptcha[12]. It is available free web services for major web development languages like ASP.NET, PHP and JSP. The reCaptcha improves the process of digitizing books by rendering a word that cannot

be read by computer or OCR technique. We consider reCaptcha as non user friendly, when it comes to reading and retyping every word (which is unknown to the users) tested. However, we believe that the user will be able to identify words for colours easily in reCaptcha (as the users are at an advantage of even predicting the word, since they expect a colour). Therefore this setup could still be considered user friendly and will have the added security.

Fig : 2 reCaptcha

II. Image based Captcha

In this type of captcha scheme the user is required to identify some image recognition task. The first image captcha was a ESP Pix and it was developed by Carneige Mellon University[13]. The given figure 3 shows the snapshot of ESP Pix captcha . In ESP Pix[14] captcha the user has provided four images and in order to pass this test , the user has to select word related to those four images from drop down list of 72 choices. In SQUIGL-PIX[15] the user traces around the border of an object in an image containing a set of objects, where the object name is presented on screen to the user. Another used image of captcha is Asirra(Animal species image recognition for restricting Access) is a cat or dog labelling captcha scheme design[16]. In this scheme, the user has to select all the pictures of cat. The given figure 4 shows the example of Asirra. The Asirra is a randomly choosing images from petfinder.com. There are many other images captcha are DIBC(Dynamic image based captcha) and other is IdentiPic, in this user has to identify the picture and each picture have drop down list having options.

Fig : 3 ESP Pix Captcha

Fig : 4 Asirra Captcha

III. Audio Captcha

In this type of captcha scheme, the user is to pass typically audio or voice recognition task. A typical audio captcha is shown in fig. 5. The general idea is to add noise and reverberation to clean speech such as existing speech recognizers can no longer recognize it. Eco[4], Byan[17] and one presented in [18] are such system.

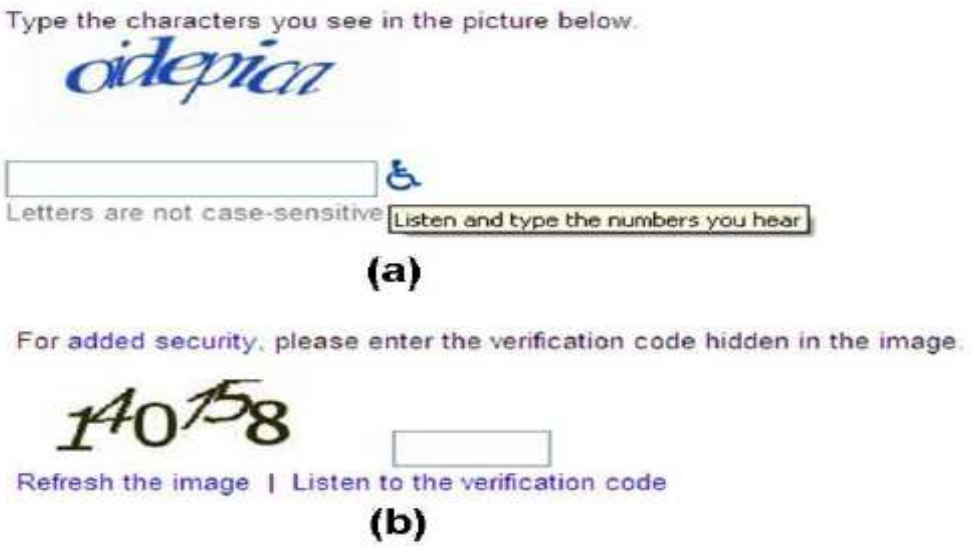

Fig : 5 Audio Captcha

IV. Line Captcha

The next captcha technique we have discussed is Line captcha. It basically gives users an image which has a distorted broken line or a set of random lines. In this scheme the user is expected to just drag the mouse along (while pressing the left button) the line continuously. When the user releases the left mouse button, the system performs the verification (also known as grading). The main objective of line captcha is its user friendliness. A user who is expected to use a line captcha will not even have to take his hand off the mouse. Just by using a "*mouse drag*" it is possible to respond to this captcha test. Random test performed with users who have no familiarity with line captchas have shown than 80% success rate in their first attempts.

V. Implicit Captcha

The captcha discussed up to now are all explicit captchas. They are introduced as an additional exercise during browsing to prove that you are human. In this method the user is asked to click on the part of the image. In 2005, Bard and Bertley proposed implicit captcha[19]. This captcha is an unconscious captcha in the form of the link on the text "MORE PHOTOS", rendered as an image. Such link text can be rendered deliberately to be difficult for machine readers, to form one in a series of effective implicit captchas. Even though implicit captchas eliminate the need for additional tasks to be performed by the users of explicit captcha, they burden the website designers to come up with innovative captchas to be embedded in their websites.

V. Collage Captcha

In this captcha scheme[20] the image of some different objects (for example four objects such as pen, kite, apple and bat) are chosen. Then some effects such as rotation are done on the images and they are merged to create a single image. This image is shown to the user and asked to click on a certain image.

VI. Captcha2

This new method [21] is like a game. Instead of decoding impossible-to-read distorted text and having to type it, need to just click two times. First click on a letter given in Captcha image proceeds the new Captcha image, click again on the correct letter complete the process.

Conclusion

Captcha is the standard for security measures on the world wide web and its applications. The captcha prevents automated scripts from online services. In this we tried to analysed all the features of captchas and carried out the systematic study on text based captcha, image based captcha and so on. We expect that this analysis will help the students or the researchers in the field of web applications security and AI. The race between makers and breakers of captchas is at a juncture where the captchas proposed today are not answerable even by humans due to their complexity and non user friendliness. We consider Captchas that are answerable by human as user-friendly. The main focus on user friendliness while not compromising the security that such systems are expected to provide.

References

[1] Luis von Ahn, Manuel Blum, Nicholas J. Hopper and John langford, "CAPTCHA: Using hard AI problems for security" Computer science Dept., Carnegie Mellon University, Pittsburgh PA 15213, USA. IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA.

[2] L Ahn, M. Blum and J. Langford. Telling Humans and Computers Apart Automatically. Communications of the ACM, 47(2):57–60, 2004.

[3] Moni Noar."Verification of a human in the loop, or Identification via the Turing test",taken from wisdom.weizmann.ac.il, Available athttp://www.wisdom.weizmann.ac.il/`nor/PAPERS/human.ps.

[4] CAPTCHA website(2000) http://www.captcha.net

[5] Bongo, http://gs9.sp.cs.cmu.edu/cgi-bin/bongo

[6] Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, Manuel Blum, reCAPTCHA: Human-Based Character Recognition via Web Security Measures, Science vol.321 Sep 2008, pp. 1465-1468.

[7] Yong Rui and Zicheng Liu, "ARTIFICIAL: Automated Reverse Turing Test using FACIAL features," Multimedia Systems, Vol 9, No.6.(! June 2004), pp. 493-502.

[8] Roman V. Yampolskiy, Venu Govindaraju. Embedded noninteractive continuous bot detection. Computers in Entertainment (CIE), vol. 5 Issue 4. March 2008. (ACM).

[9] P. Golle and N. Ducheneaut. Preventing bots from playing online games. Comput. Entertain., 3(3):3, 2005.

[10] Hilaire, S.; Hyun-chul Kim; Chong-kwon Kim; How to deal with bot scum in MMORPGs?; IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR), 2010. Pp: 1 – 6.

[11] Benny Pinkas and Tomas Sander: Securing Passwords against dictionary attacks. ACM Conference on Computer and Communications Security 2002: 161-170.

[12] reCAPTCHA project, available at http://recaptcha.net.

[13] ESP Pix CAPTCHA, taken from cmu server, Available at http://server251.theory.cs.cmu.edu/cgi-bin/esp-pix/esp-pix.

[14] ESP-PIX,[Online]. Available: http://www.captcha.net/cgi-bin/esp-pix. [Apr.10,2009]

[15] SQUIGL-PIX, [Online]. Available: http://www.toallwhoseekit.net/cgi-bin/sq-pix. [Apr.10,2009].

[16] Captcha the dog, available at, http://www.captchathedog.com.

[17] Chen, N., Byan, http://drive.to/research.

[18] Kochanski, G. et al(2002): "A Reverse Turing Test Using Speech", In: Proceedings of the Seventh International Conference on Spoken Language Processing (ICSLP 2002 - INTERSPEECH 2002), Denver, Colorado, USA, September 16-20, 2002, 1357-1360.

[19] Henry S. Baird and Jon L. Bentley, Implicit CAPTCHAs, Proceedings, IS&T/SPIE Document Recognition & Retrieval XII Conference, San Jose, CA, January 16-20,2005.

[20] M. Shirali-Shahreza and S. Shirali-Shahreza, "Collage CAPTCHA." In Proceeding of the 20th IEEE International Symposium Signal Processing and Application (ISSPA 2007), Sharjah, United Arab Emirates (UAE), February 2007.

[21] http://Captcha2.com/