Volume 3, Issue 11, November 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

CAPTCHA Based Web Security: An Overview

Sushama Kulkarni*
Research Scholar,
Department of Computational Science,
S. R. T. M. University, Nanded, Maharashtra, India.

Dr. H. S. Fadewar
Assistant Professor,
Department of Computational Science,
S. R. T. M. University, Nanded, Maharashtra, India.

Abstract— Completely Automatic Public Turing test to tell Computers and Humans Apart (CAPTCHA) is a HIP (Human interactive Proof) system. CAPTCHAs are used to improve the security of Internet based applications in order to ensure that a web based application which is intended to be used by a human being is not maliciously used by Artificially Intelligent programs called bots. As the current CAPTCHA methods are striving to turn out to be difficult for bots, they are gradually becoming difficult and annoying for human users as well. In this paper, we review the existing CAPTCHA schemes and the trend of using AI-Complete problems for designing efficient CAPTCHAs.

Keywords— Completely Automatic Public Turing test to tell Computers and Humans Apart (CAPTCHA), Human Interactive Proof (HIP), Turing Test, Web Security, AI-Complete

I. INTRODUCTION

CAPTCHA (Completely Automatic Public Turing Test to Tell Computer and Human Apart) is a Human Interactive Proof (HIP) system which is used to distinguish between human users and computer programs automatically [1]. The thumb rule of CAPTCHA is that it should be solved easily by a human but not by a bot.

Sometimes CAPTCHA is also referred as Reverse Turing test. It has following specifications:

- The judge is a machine instead of a human.
- The goal is that virtually all human users will be recognized and pass the test, whereas no computer program will pass.

CAPTCHA helps to prevent artificially intelligent automated software programs known as bots (that pose as human users) from performing malicious activities like spamming and other fraudulent activities. These Web-bots pose a major threat to web services. Web-bots try to automatically register for a large number of free accounts and then use these accounts to spam legitimate users by sending junk e-mail messages or slowing down the service by repeatedly signing on to multiple accounts simultaneously or causing other denial of services. Hence most of the websites have adopted CAPTCHA as defensive scheme against such Web-bots. Its methods are based on Artificial Intelligence (AI). If a CAPTCHA can be solved programmatically it marks scientific progress on a hard AI problem [3]. A problem which cannot be solved by computer programs can be used as CAPTCHA. The existing text-based CAPTCHAs are not safe as computer-vision techniques develop rapidly. The modified solutions are still in the stage of infancy because they are either hard to solve and costly to implement or easily trespassed due to compromised security. This indicates that continuous efforts are required to improve the robustness of CAPTCHA.

II. RESEARCH WORK CONDUCTED IN OCR BASED CAPTCHA

OCR-based CAPTCHAs are mainly text-based CAPTCHAs in which the user is shown distorted images of letters and/or digits and the user is required to recognize them and type the answer. But, these CAPTCHAs have an inbuilt drawback. The strength of OCR-based CAPTCHAs extensively depends upon the degree of distortion in the displayed text and if increasing security is achieved by increasing text distortion, it may lead to failure of recognition by humans, thus making the CAPTCHA ineffective. In addition, OCR-based CAPTCHAs are problematic for mobile phones and devices like PDAs and palmtops, as the use of keyboard may be infeasible or difficult.

Some of the OCR-based CAPTCHA methods are reviewed below:

A. Text-based CAPTCHA

Among visual CAPTCHAs, Text-based CAPTCHA is one of the most popular types. It exploits the ability of people to read images of text more reliably than Optical Character Recognition (OCR) or other machine vision system. As these CAPTCHAs are becoming more difficult for genuine users, attackers are also getting better at breaking existing CAPTCHAs [12].

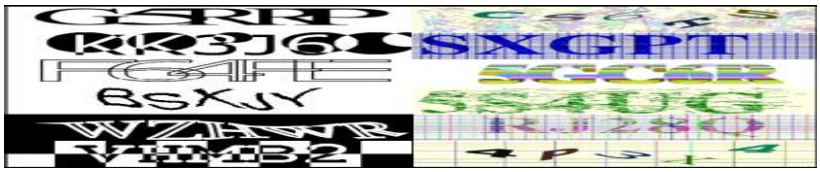

Fig.1 Text-based CAPTCHA

© 2013, IJARCSSE All Rights Reserved

B. Gimpy method

As this method uses its word from a dictionary with 850 words, it can easily be broken in. A correlation algorithm was developed that correctly identified the word in EZ-Gimpy CAPTCHA 99% of the time and a direct distortion estimation algorithm that correctly identifies the four letters in a Gimpy-r CAPTCHA 78% of time [13].

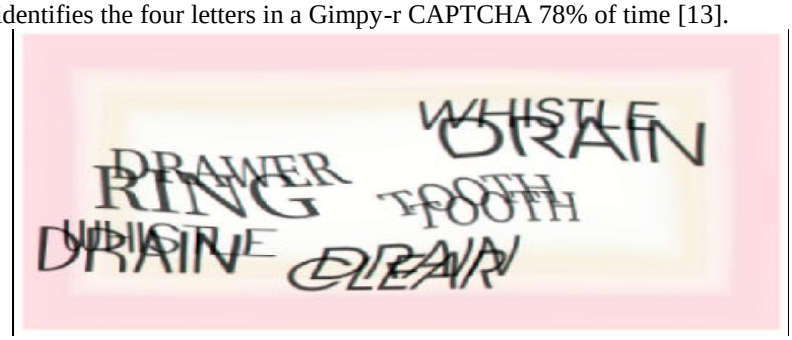

Fig.2 Gimpy CAPTCHA

C. Pessimal Print Method

This method tries to prevent the operations of destructive computer software by artificially lowering the quality of the printed letters [5]. Two main methods of attack on Pessimal Print are using the Mori-Malik algorithms and brute-force.

Fig.3 Pessimal Print CAPTCHA

D. Baffletext Method

In the Baffletext method, words that are not provided in English dictionaries are produced, and then the picture of the word is changed with different degrees of ease or difficulty [14]. These text-based CAPTCHAs are prone to bot attacks.

Fig.4 Baffletext CAPTCHA

III. RESEARCH WORK CONDUCTED IN NON-OCR-BASED CAPTCHA

Non-OCR based CAPTCHAs basically test the audio/video sense capability of a human being. Some of these methods are reviewed below.

A. Implicit CAPTCHA

In the Implicit CAPTCHA, users only have to make a simple click [6]. For example, in this method, the picture of a mountain is shown to the user and the user is asked to click on the pinnacle of the mountain.

B. Audio CAPTCHA

In this method, instead of showing an image, a sound is played which the user must recognize and type the word. Researchers succeeded in breaking three different types of widely used audio CAPTCHAs with 71% accuracy [15].

C. Video CAPTCHA

In Video CAPTCHA, a user has to provide appropriate tag for the video displayed as a CAPTCHA test [16]. It is not easy for bot as compared to human being. But they take more time to get loaded on a web page which may turn of a genuine user. Video CAPTCHA are prone to bot attacks which use database replication, Video analysis, etc.

IV. RESEARCH WORK CONDUCTED IN COGNITIVE CAPTCHA

Cognitive CAPTCHAs are those which use human cognitive skills like classification, grouping, interpretation, game playing, etc. for preventing bots. In fact, cognitive CAPTCHAs use AI-hard or AI-Complete problems to identify humans and bots apart. Some of the cognitive CAPTCHAs are reviewed below:

© 2013, IJARCSSE All Rights Reserved

A. Question-based CAPTCHA

Proposed in 2009, this CAPTCHA explored various skills of a user through a question which can only be answered by a human user [17]. Later a dynamic question based CAPTCHA was also proposed.

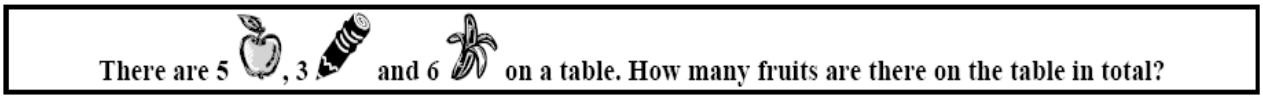

Fig.5 Question-based CAPTCHA

B. Math CAPTCHA

This CAPTCHA asks user to solve a mathematical equation in order to pass the test [18]. Difficulty level of the equation varies based on various implementations.

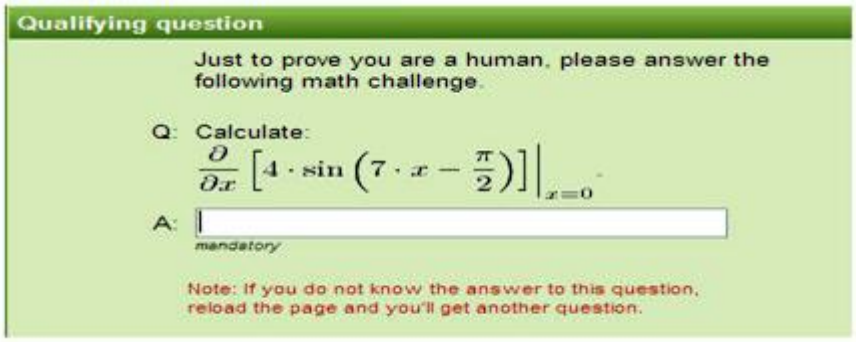

Fig.6 Math CAPTCHA

C. NLP CAPTCHA

As the name suggests, this CAPTCHA method relies on human capability of Natural Language Processing. It efficiently makes use of advertisements which are embedded with the challenge for users [19].

Fig.7 NLP CAPTCHA

D. Game CAPTCHA

This CAPTCHA method uses a database of cartoon mini-games that are interesting and supportive for users with accessibility difficulties as well. These CAPTCHAs are not just suitable for desktop, but they are adapted for mobile and touch-screen devices also [20].

Fig.8 Game CAPTCHA

V. CAPTCHA FOR WEB SECURITY

The official CAPTCHA website enlists several applications of CAPTCHA test for practical security including (but not limited to):

- Preventing Comment Spam in Blogs.
- Protecting Website Registration.

© 2013, IJARCSSE All Rights Reserved

- Protecting email addresses from scrapers and Search engine bots.
- Online polls.
- Preventing Dictionary Attacks and Worms [23].

Some strategies were suggested to improve the strengthening CAPTCHA based web security. This concluded that presenting multiple CAPTCHA systems together in random order may provide quantitative and qualitative advantages over many typical present-day CAPTCHA systems.

A study suggested that CAPTCHA implementation should employ Global Unique Identifier (GUID) to ensure that sender of CAPTCHA solution is really the computer which was send a CAPTCHA challenge by the server [7]. Critical vulnerabilities in various text based CAPTCHA schemes were identified from security engineering perspective [9].

VI. USABILITY OF CAPTCHA

Only a few works studied the usability issues of CAPTCHAs. A study discussed variety of factors that should be considered when designing CAPTCHAs [8]. Another research illuminated the factors affecting the readability of CAPTCHA designs [10]. Recently, researcher evaluated impact of colour on usability of text based CAPTCHA [8]. The just mentioned works evaluated the usability of CAPTCHA. It is observed that overall usability of CAPTCHA decreases with increase in complexity.

VII. AI-COMPLETENESS OF CAPTCHA

In 1950 Alan Turing published his best known paper "Computing Machinery and Intelligence" in which he proposed evaluating abilities of an artificially intelligent machine based on how closely it can mimic human behaviour [22]. The test, which is now commonly known as the Turing test, was in the form of conversation conducted by a human evaluator for two agents present in separate room, one of them would be a human and other being the machine. Based on the conversation, human evaluator would identify the human and machine apart. Now there is a similar requirement of identifying human and artificially intelligent bots apart on the web in order to maintain the security. But employing human evaluator for identifying millions of web users is not a feasible solution. Hence there was a need of automating the classic Turing test and thus theoretical platform for an automated Turing test (ATT) was developed by Naor in 1996, where he replaced human evaluator with a computer [21]. In addition to ATT, the developed procedures are known under such names as: reverse Turing test (RTT), human interactive proof (HIP) or completely automated public Turing test to tell computers and humans apart (CAPTCHA) [3]. As such, the term "AI-Complete" (or sometimes AI-Hard) has been a part of the AI field for many years and has been frequently brought up to express difficulty of a specific problem investigated by researchers. Turing test problem is an instance of an AI-complete problem. A research emphasized the use of such AI-hard or AI-Complete problems while designing CAPTCHA for web security [2]. A study explored the space of systematic distortions as a means of making automated image matching and recognition a very hard AI problem used for designing image based CAPTCHA [4]. Amalia Rusu used cognitive aspects like perception, context, syntax and reading comprehension of auto-generated handwritten text for designing CAPTCHA for web security applications [11].

VIII. CONCLUSION

Various CAPTCHA alternatives are continuously emerging, and this race will continue as more advanced bots emerge. However, the basic idea of CAPTCHAs is to tell humans and machines apart, and this concept is still worth to be discovered for several reasons. CAPTCHAs today are making use of AI-hard problems such as cognitive skills of human being for preventing bots and thus ensuring the security of web applications. Future trends in CAPTCHA techniques henceforth need to encourage the application of AI-hard problems for efficient prevention of bots.

REFERENCES

[1] L. von Ahn, M. Blum and J. Langford, "Telling Humans and Computer Apart Automatically," in *Communications of the ACM*, vol.47, no. 2, pp. 57-60, 2004.

[2] L. von Ahn, M. Blum, N.J. Hopper and J. Langford, "CAPTCHA: Using hard AI problems for security," in *Proc. of Eurocrypt '03*, pp. 294-311.

[3] H.S. Baird and K. Popat, "Human Interactive Proofs and Document Image Analysis," in *Proc. of the 5th IAPR International Workshop on Document Analysis Systems*, Springer LNCS 2423, pp. 507-518, 2002.

[4] Datta, R., Jia Li, Wang, J.Z., "Exploiting the Human–Machine Gap in Image Recognition for Designing CAPTCHAs," *IEEE Transactions on Information Forensics and Security*, vol.4, no.3, pp.504-518, 2009.

[5] A.L. Coates et al, "Pessimal Print: A Reverse Turing Test," in *Proc. of the 6th International Conference on Document Analysis and Recognition*, Seattle, WA, USA, pp. 1154-1158, 2001.

[6] H.S. Baird and J.L. Bentley, "Implicit CAPTCHAs," in *Proc. of SPIE/IS&T Conference on Document Recognition and Retrieval XII (DR&R2005)*, San Jose, pp. 191-196, 2005.

[7] M. Tariq Banday and N. A. Shah, "A Study of CAPTCHAs for Securing Web Services," In *IJSDIA International Journal of Secure Digital Information Age*, Vol. 1. No. 2, pp. 101-105, December 2009.

[8] Ahmad El Ahmad, Jeff Yan, Wai-Yin Ng, "CAPTCHA Design: Color, Usability, and Security," *IEEE Internet Computing*, vol. 16, no. 2, pp. 44-51, 2012.

[9] Jeff Yan, Ahmad Salah El Ahmad, "Captcha Robustness: A Security Engineering Perspective," *IEEE Computer*, vol. 44, no. 2, pp. 54-60, 2011.

© 2013, IJARCSSE All Rights Reserved

[10] Aleksey Kolupaev, Juriy Ogijenko, "CAPTCHAs: Humans vs. Bots," *IEEE Security & Privacy*, vol. 6, no. 1, pp. 68-70, 2008.
[11] A. Rusu., "Exploiting the Gap in Human and Machine Abilities in Handwriting Recognition for Web Security Applications," Ph.D. thesis, Dept. of Comp. Sci. and Eng., University of New York at Buffalo, USA, August 2006.
[12] Kumar Chellapilla, Kevin Larson, Patrice Y. Simard, and Mary Czerwinski, "Building Segmentation Based Human-Friendly Human Interaction Proofs (HIPs)," in *Proc. of HIP 2005*, pp. 1-26, 2005.
[13] Gabriel Moy, Nathan Jones, Curt Harkless, and Randall Potter, "Distortion Estimation Techniques in Solving Visual CAPTCHAs," in *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, vol. 2, pp. 23-28, 2004.
[14] Chew M. and Baird H. S., "BaffleText: a Human Interactive Proof," in *Proc. of 10th SPIE/IS&T Document Recognition and Retrieval Conference (DRR2003)*, Santa Clara, CA, USA, pp. 305-316, 2003.
[15] Tam J., Simsa J., Hyde S. and Von Ahn, "Breaking Audio CAPTCHAs," *Advances in Neural Information Processing Systems*, 2008.
[16] K.A. Kluever and R. Zanibbi., "Balancing usability and security in a video CAPTCHA," in *Proc. of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*, ACM, New York, NY, USA, Article 14, pp. 1-11, 2009.
[17] Mohammad Shirali-Shahreza and Sajad Shirali-Shahreza, "Question-Based CAPTCHA," in *Proc. of International Conference on Computational Intelligence and Multimedia Applications*, pp. 54-58, 2007.
[18] C. J. Hernandez-Castro and A. Ribagorda, "Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study," *Computers & Security*, vol. 29, no. 1, pp. 141-157, 2010.
[19] http://nlpcaptcha.in/.
[20] http://areyouahuman.com/.
[21] Naor Moni, "Verification of a human in the loop or identification via the turing test," http://www.wisdom.weizmann.ac.il/~naor/PAPERS/human.ps, 1996.
[22] Turing, A. M., "Computing Machinery and Intelligence," *Mind*, vol. 59, no. 236, pp. 433-460, 1950.
[23] http://www.captcha.net

© 2013, IJARCSSE All Rights Reserved