Abstract

A CAPTCHA is a widely used security mechanism to prevent automated access to webpages by malicious users such as internet bots. Introducing elements such as distortion, distracting backgrounds and noise are some ways to make it harder for computer vision algorithms to break them. However, making CAPTCHAs complex by increasing the intensity of these properties has a direct adverse impact on their legibility to humans and hence usability. We conduct an eye tracking study on participants to understand how ease of comprehension of CAPTCHAs varies across increasing levels of complexity. We seek to find out where to draw the line in increasing complexity for CAPTCHA generation, in order to present ‘good’ CAPTCHAs on websites that are balanced in terms of their usability and security against internet bots.

CR Categories: H.1.2 [User/Machine Systems]: Human Factors, H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/Methodology

Keywords: CAPTCHA usability, CAPTCHA complexity, eye tracking

1 Introduction

CAPTCHA is an acronym for ‘Completely Automated Public Turing Test to tell Computers and Humans Apart’. With the tremendous popularity of the internet and its practically irreplaceable role in almost all of our activities, securing the web is an important issue. The internet is susceptible to a large variety of threats that can compromise any system in the absence of secure applications providing protection against such threats. One such threat is an internet bot, a program capable of running malicious automated tasks and creating problems in the network [1]. It can abuse web facilities by activities such as creating junk accounts and leaving spam comments on websites. A CAPTCHA is one mechanism that can provide protection from such malicious programs.

There is a variety of CAPTCHAs such as audio, video, puzzle/question based which are in use today. The most popular ones are still text based CAPTCHAs. These present an image of some text which is made complex and noisy in various ways. The users are asked to type in what they see. The objective is to make it difficult for a machine to pass this test. CAPTCHAs are intended to be easy for humans and difficult for machines to decipher. However, with advances in machine vision it is becoming easier to develop efficient character recognition systems that can break these mechanisms. As a result, the complexity of CAPTCHAs being presented on websites has been increasing consistently. This presents a new challenge - it becomes much harder for human users to comprehend CAPTCHAs as well. We seek to address this issue by studying, using eye tracking, how users perceive CAPTCHA complexity. Our goal is to set a threshold on complexity for a specific class of CAPTCHAs – distorted text CAPTCHAs. We seek to estimate the level beyond which CAPTCHAs become illegible to humans or even hard enough that putting efforts in reading them leads to frustration and poor accuracy.

2 Background

A pilot study was conducted by Hend S. Al-Khalifa [2] that used eye tracking on participants reading CAPTCHAs. However, they only discuss empirically that data from eye tracking metrics (fixation count and duration) does relate to complexity of CAPTCHAs as perceived by users. We evaluate their usability at varying quantifiable levels of complexity and estimate an optimum level that generates most usable CAPTCHAs.

Chellapilla et al. [3] conducted a user study to examine the impact of common techniques used in CAPTCHA design, including distortion. In their study they calculated users’ accuracy in solving CAPTCHAs to measure usability under different distortion settings. However, evaluation based on accuracy goes only so far as being able to indicate when CAPTCHAs become illegible, but not cumbersome. Our approach to study the impact of distortion on CAPTCHA usability using eye tracking addresses this issue.

Fidas et al. [4] used a questionnaire-based survey to determine end-user perceptions about the usability of CAPTCHAs. Eye tracking as a technique is impressive in its efficacy in usability analysis as it reveals usage patterns precisely. This is especially useful in a study like ours where we seek to find out precisely the degree to which complexity of CAPTCHAs can be increased before making them too difficult to read. Eye tracking is a fairly good alternative to questionnaire-based surveys for the purpose of usability analysis on CAPTCHAs.

3 Methodology

We conduct an eye tracking experiment with participants to study their eye tracking data as they read out CAPTCHAs. Next, we present the setup and procedures of our experiment.

3.1 Apparatus

The eye tracker that we used was the Gazepoint GP-3 Desktop eye tracker from Gazepoint (figure 1). The Gazepoint is a video based pupil-corneal reflection eye tracker operating at 60Hz with 0.5-1 degree of visual accuracy. The eye tracker offers 5 point and 9 point calibration and horizontal - vertical movement of 25 cm X 11 cm. The eye tracker was mounted on and used with a desktop PC with 17 inches monitor operating on Windows 7 (64bit). An in-built microphone of a mobile device was used for audio recording participants as they read aloud the CAPTCHAs.
3.2 Stimulus

The participants were presented with CAPTCHAs (figure 2) at five different levels of complexity in random order. The scale of complexity was from 1 to 5, with 1 being the least complex and 5 being the most complex. The complexity was altered by changing two main factors - distortion and overlap. We processed 10 unique CAPTCHAs, 2 for each level of complexity, composed of random letters. To embed overlap, spacing in text was condensed using Microsoft Word from 2 point to 6 points with increment of 1 point at each level. Font style used was Gabriola at size 72. Next, the text was clipped from MS Word and the images were processed in Adobe Photoshop to produce distortion. Filters used to produce distortion were wave, pinch and twirl in the same order.

The following characteristics were used for filters in CAPTCHA generation –

- **Wave** - Generators, with minimum wavelength of 10 units and maximum wavelength of 120 units, minimum amplitude of 5 and maximum amplitude 35 units. Horizontal and vertical scale of 35% to 55% with increments of 5% at each level
- **Pinch** – Negative 65% to Negative 85% with decrements of 5% at each level
- **Twirl** - 55 units to 75 units in increments of 5 units at each level

3.3 Subjects

Ten participants, of ages between 18 and 28 were recruited for the study. Participants were a mix of undergraduate and graduate students of Clemson University. All the participants were either native speakers or had full professional proficiency in the English language. All the participants were frequent users of the web and had encountered and solved CAPTCHAs in the past.

3.4 Experimental Design

We use the within-subjects approach with our experiment. The factor manipulated in the experiment is Complexity which is varied across 5 levels. Each level of complexity is tested with 2 CAPTCHAs. The agent introducing complexity is single dimensional, even though we clearly identify two properties namely overlap and distortion that increase complexity. This is justified, as we operate with the combination of the two and at each level of complexity, both the parameters are increased simultaneously. Each participant is thus presented with 5x2x1, that is, 10 CAPTCHAs.

We analyze the eye tracking data with 4 metrics – fixation duration, fixation count, revisit count and accuracy of responses. Our hypotheses with respect to these metrics are as under.

H1: CAPTCHAs belonging to higher complexity levels will produce higher average fixation duration than those belonging to lower levels.

H2: CAPTCHAs belonging to higher complexity levels will produce higher average fixation count than those belonging to lower levels.

H3: CAPTCHAs belonging to higher complexity levels will produce higher average revisit count than those belonging to lower levels.

H4: CAPTCHAs belonging to higher complexity levels will produce equal or lower accuracy rate of responses than those belonging to lower levels.

3.5 Procedure

Participants were asked to sign a consent waiver form before participating in the experiment. After accepting to participate in the experiment, the participants were asked to fill up a short questionnaire on the participant’s basic demographic information and web
usage details. Participants were then asked to sit at a distance of around 50 cm from the display monitor and get familiar with the eye tracking equipment. The eye tracker was then calibrated for each participant using nine calibration points. After the calibration was completed, the participants were asked to begin the test when they were ready.

The CAPTCHAs were shown to the participants on ‘Gazepoint Analysis’ application and the participants were able to change the CAPTCHAs by pressing the space bar. The participants were asked to read each CAPTCHA aloud. This ensured minimal navigation and distraction of eye movements due to supportive tasks. The participants were audio recorded as they read aloud the CAPTCHAs. The audio recordings were used to calculate the accuracy of their responses.

4 Results

The metrics we use to evaluate CAPTCHA complexity from the captured eye tracking data were compiled by the Gazepoint Analysis program.

From figure 3 and 4, we can visualize that there are more number of fixation points in the level 4 CAPTCHA than there are in level 3, leading to a higher fixation count. Also, we observe that the size of fixation points is larger in the level 4 CAPTCHA than in level 3, indicating larger fixation duration. The overall results for these metrics are summarized in plots in Fig. 6 and Fig. 7. From Fig. 5, we can observe several saccadic traces going back and forth within the marked region. An AOI to the left of the marked region produced several revisits. If we compare Fig. 5 with Fig. 4 we can clearly visualize that Level 5 produces much more revisits than level 4. The overall results for revisits is summarized in Fig. 8. Results for accuracy measure which were evaluated with the help of audio recordings of participants are summarized in Fig. 9. These verify our hypothesis H1 through H4.
5 Discussion

Eye-tracking data provides much better insight on CAPTCHA complexity than other methods like purely accuracy based analysis or questionnaire approaches. The eye tracking data helps us define a threshold for generating CAPTCHAs which are both secure and easy to read.

In our study, the CAPTCHAs with complexity level 5 were the most difficult to read for the users and an accuracy rate of 60% was obtained at this complexity level. Whereas, the captchas with complexity level 1 and level 2 were the easiest for the users to read and all the participants had a 100% accuracy while reading these captchas. However, level 2 CAPTCHAs still produce higher values of eye tracking metrics than those belonging to level 1. Level 3 had a good accuracy rate of 90%, so did Level 3 at 85%. However, the values of eye tracking metrics for CAPTCHAs at complexity level 4 are close to those at level 5 whose accuracy rate is only 60%. Also, we see that there is a fairly sharp increase in the number of fixations, fixation duration and the number of revisits between level 3 and level 4. This gives us an indication that CAPTCHAs with complexity level 4, despite having a good accuracy rate, are still difficult for the participants to read. In other words, participants were not nearly as confident while reading level 4 CAPTCHAs as they were while reading level 3 CAPTCHAs.

Thus we can say that complexity level 3 an optimum level of CAPTCHA complexity out of all the five complexity levels in terms of usability.

6 Conclusion

Eye tracking as a technique can provide insights into the complexity of CAPTCHAs as perceived by end users more precisely and at a finer scale than other techniques. Combining techniques such as accuracy evaluation with eye tracking and analyzing their interaction can provide a precise measure of perceived CAPTCHA complexity. This is especially helpful in knowing the relative perceived complexities of two CAPTCHAs of slightly different theoretical complexities where accuracy measures are almost consistent. In reality, users will perceive such CAPTCHAs with different confidence level even though they may end up answering both of these CAPTCHAs correctly. Eye tracking metrics are a good means to indicate this confidence level or perceived complexity in CAPTCHAs.

References