Implicit Drag and Drop Image CAPTCHA for Web Security

Adebayo Omotosho, Ukeme Asanga, Emmanuel O. Asani, Joyce Ayoola and Paula Fiddi

Department of Computer Science, Landmark University, Omu-Aran, Nigeria

Department of Computer Science and Information Technology, Eells University of Technology, Ota, Nigeria

Department of Computer Science, University of Oxford, Oxford, United Kingdom

omotosho.adebayo@lmu.edu.ng

Abstract: This study proposes an implicit drag and drop image Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) that can be used in distinguishing human users from web bots. The presented theoretical scheme will mitigate one of the problems associated with regular text-based CAPTCHA which is the challenge of identifying text by users with low vision. The 2D images of basic shapes, rather than text in a drag and drop CAPTCHA Mode were proposed to be presented to users and the relative distance of the displayed images to the drop boxes is then used to manipulate the bits of confidence property against a random clicker or bot. By following this approach, this research is expected to ease CAPTCHA usage and improve security of online systems as well as reduce automatic access from web bots.

Key words: CAPTCHA, security, web application, bots, random clicker, drop boxes

INTRODUCTION

Over the years, the number of internet users continue to grow exponentially, according to the International Telecommunication Union, there are about 3.9 billion projected users as at 2016 but many of them are malicious and originating from automated computer program resulting in different negative consequences on security, healthcare, economy and individuals (ITU, 2016). A report by PWC, (2017) showed that the number of cyber security breaches keep increasing at accelerated pace with a remarkably high increase of 38% in 2015 as compared to 2014. The sophistication of modern bots represent a serious challenge and it has become more difficult for CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) systems to differentiate humans from non-human users (Madar et al., 2017; Singh and Jasmine, 2017; Kumar and Ramachandran, 2017; Hernandez-Castro et al., 2017). Threats from bots automated scripts are growing geometrically with their attendant devastating effect on security and privacy of legitimate human users. For instance, it is possible for legitimate users to be deceived, denied and cheated by false requests on their online and offline social life (Ferrara et al., 2016; Marechal, 2016; Yang et al., 2016; Stiegitz et al., 2017; Tracre et al., 2017).

CAPTCHAs are generally used to block automated access in online applications or services. Most of the existing text-based CAPTCHA systems are susceptible to Optical Character Reader (OCR) which are capable of recognizing printed or written text characters by a computer through photo-scanning of the text character-by-character, analysis of the scanned-in image and then translation of the character image into character codes such as American Standard Code for Information Interchange (ASCII), commonly used in data processing (Von Ahn et al., 2008; Sami and Bala, 2013; Azad and Jain, 2013; Singh and Pal, 2014). Two flavours of text-based CAPTCHA are shown in Fig. 1 and 2. Text based CAPTCHAs have also received some criticisms for the fact that they are language (mostly English) dependent and that visually impaired users may have difficulties recognizing distorted text (Nanglae and Bhattachakosol, 2015; Aldosari and Al-Daraiseh, 2016). Audio based CAPTCHAs are available in English, so, user must have a comprehensive English vocabulary. It is also, problematic differentiating characters that have similar sounds. Video based CAPTCHA have to grapple with the challenge of space due to the sizes of the files involved (Nanglae and Bhattachakosol, 2015; Ye et al., 2013). There is a need to develop a better, simple and widely acceptable approach that will help to overcome the challenges in existing CAPTCHA system and improve security of online resources which are vital to organizations, individuals and government. This study proposes an implicit drag and drop image CAPTCHA that
can be used in distinguishing human users from web bots. It is implicit in the sense that the CAPTCHA is not communicated directly but can be implied indirectly. As multiple shapes can be presented, users experience and knowledge of shapes identification will help in distinguishing and solving the CAPTCHA. Images in the form of basic shapes will be used as CAPTCHA image. The proposed scheme will be suitable for people with low vision because it is language independent.

**Literature review:** Some of the selected literature are summarized chronologically as follows, CAPTCHA as graphical password has been implemented with other media besides image-only. Saranya et al. (2016) suggested the use of image and audio media in hard AI problems CAPTCHA authentication. The purpose of their research was to improve the security for any website that depends on graphical password with persuasive cue click points. The researchers incorporated sound signature in graphical password system in order to increase password memorability. Pass point scheme was used that comprises several points anywhere on an image. Also, only login attempts that were approximately correct were accepted. The researchers only provided a brief theoretical description but the system was not implemented.

Zhu et al. (2014) developed CAPTCHA as graphical passwords (CaRP) which integrates both CAPTCHA and graphical authentication based on hard AI problems. CaRP caters for a number of security challenges such as online guessing attacks and relay attacks. CaRP’s design incorporates both text-based CAPTCHA and image-recognition CAPTCHA. CaRP is click-based, the graphical passwords is generated after a sequence of clicks on an image. As opposed to similar click-based graphical passwords, images used in CaRP also serve as CAPTCHA challenges, thus, the system pops up a new CaRP image for every login attempt to make trials of an online guessing attack computationally independent of each other. A password of CaRP can be determined only stochastically by automatic online guessing attacks and brute-force attacks. This security feature is lacking in other graphical password schemes.

Deepika and Babu (2015) also proposed the use of CAPTCHA as graphical password (CaRP) based on hard AI problems. As noted by Zhu et al. (2014), the merit of CaRP is that password cannot be guessed easily, attack vector based automatic online guessing attacks such as brute-force attacks have to rely on probability to obtain the password. The technique presented by Deepika and Babu (2015) is similar to Davis et al. (2015). Survey result on ease of use of the implemented system was encouraging. More users adjudged AnimalGrid and ClickText easier to use than either PassPoints or a combination of text password and CAPTCHA. AnimalGrid and ClickText exhibit better password memorability than the conventional text passwords. The researchers recommended CaRP’s ease of use can be further enhanced by popping images in increasing levels of difficulty based on the login history of the user and the machine used to log in. They concluded that additional works need to be done to improve the optimal tradeoff between the security and usability of CaRP. Kumesh and Rani (2015) also reported the deployment of a very similar CaRP system.

Davis et al. (2015) and Thrivikram et al. (2015) presented a theoretical description of a CAPTCHA as graphical passwords (CaRP) that relied on unsolved hard AI problems to circumvent the problem of online guessing attacks. They conceptualized the use of distinctive CaRP image for each login CAPTCHA challenge, thus, making online guessing attack computationally difficult and isolated of each other. The overall system design forces attackers to resort to significantly less efficient, less cost efficient and more error prone human-based attacks. The researchers proposed that future research should improve the login time and memorability. The approach was not implemented and there was no justification for the use of small threshold and large threshold for failed attempts from an unknown and known machine, respectively.
Kumar and Sasikala (2014) presented a technique for secure online activities using mutual authentication and Clicking-Cropping based image CAPTCHA technology. The system has registration and login modules. The sign-up process for first time users includes users providing unique username and password. The system then pops a set of images from which the user is required to pick an image for the crop and click registration. The position of the cropped image on the (x, y) cartesian coordinates and the number of clicks on that image is then stored in the database as part of the user’s authentication record. The user must then supply the necessary authentication details, that is the username and password at login. After this, the user is required to select an image from a set of images and crop it. Login will be unsuccessful until the user input matches the number of clicks in the stored record. The system demonstration was done with windows application modelled after the searches of Gao et al. (2010).

The image-based CAPTCHA developed by Gao et al. (2010) involves solving a jigsaw puzzle of an image segments arranged in a (n by n) matrix. Two of the segments are misplaced. The challenge is to solve the puzzle by re-arranging the misplaced segments in the right quadrant. In order to limit computer’s abilities to solve the puzzle, the edge is modified to avoid edge detection. Similar to Goswami et al. (2014), there was no text entry required and it was language independent. A survey to determine the ease of use of the system was conducted via email to 100 respondents. Respondents, majority of whom are college students tested the system with the online version of the developed windows application. The puzzle to solve the 3 by 3 image matrix was 89% successful, the 4 by 4 image matrix was 88.7% successful rate of 88.7% while the 5 by 5 image matrix was 87.6%. Consequently, we may infer that the difficulty increases with the dimension of the image matrix. It was also, reported that 88 of respondents would rather complete the jigsaw puzzle that use text-based.

The research by Goswami et al. (2014) deviated from language dependent CAPTCHA to propose a face detection CAPTCHA called FaceDCAPTCHA. In this study, users are to complete a CAPTCHA challenge by identifying genuine face and marking its approximate center from a selection of distorted genuine and fake images places on random background. The challenge is adjudged solved, if the user identifies and mark the genuine image as appropriate, if otherwise it fails. The hypothesis is that it is more difficult for bots to solve the carefully crafted challenge whereas humans can solve it with ease. The major benefits of this approach over existing methods is that it optimized for mobile devices because of its language independence and the fact that it can use alternative means for data entry. By Nejati et al. (2014) proposed an image-based CAPTCHA called DeepCAPTCHA to circumvent limitation inherent in text-based CAPTCHAs. The system takes advantage of the human ability of depth perception and ability to reliably and quickly identify and compare objects to solve DeepCAPTCHA while machines are having difficulties completing these challenges. Users were expected to arrange in terms of size or depth, 3D objects (such as animals, plants, furniture and so on) that are loaded from a 3D Model dataset created from Internet sources using web crawlers. Each model in the dataset is anonymized by renaming it and removing all identifying attributes as tags and object metadata. A rough machine learning classification was carried out on each new model to determine its difficulty for machines to learn the object’s signature. Merge Sort algorithm was then used to order these unknown objects. Experimental metrics for human’s accuracy in solving the DeepCAPTCHA challenge was (~8.4%) while machines performed poorly.

Aadhirai et al. (2012) developed a CAPTCHA system that relies on relative distances based on the information provided by a two-dimensional picture to distinguish humans from bots. Through emphasis on perception and edge interpretation, humans must use their ability to understand relative distances and sizes of objects from the point of perception in an image to answer the challenge presented by the system. This is based on the fact that identification based on distances are conveniently obvious to humans than computers. The CAPTCHA scheme was tested with a predefined and customized database of 75 images and queries instead of dynamic scene-query generation. For the system evaluation, 50 participants from a wide range of professions and computer literacy volunteered to answer the CAPTCHA challenge. The user experience is rated as 9 on 10 by the participants of the controlled test run and a success rate of 94% was obtained which is a high value. However, automated attack testing was not carried out on the system to ascertain its vulnerability. The researchers proposed that future research should include automating the generation of images required which will overcome the drawback of sorting the existing images for the implementation. Also, an algorithm may be developed in order to reduce the manual work involved in choosing the right image.

The concept building an image (graphic) CAPTCHA using biological (human) and non-biological (virtual world avatar) faces motivated the research by D’Souza et al. (2012) on avatar-based CAPTCHA through face classification. The proposed CAPTCHA asks users to
identify avatar faces from a set of 2 rows with 6 grayscale images each that comprised a mix of human and avatar faces randomly picked from a developed dataset. Each image has a checkbox associated with it for the user to make his choice. The use of grayscale images is to prevent computer programs from breaking the CAPTCHA by taking advantage of the varying color spectrum difference between human and avatar images. The image datasets used were obtained in real-time and comprise human and avatar images from popular online websites such as Flickr and ActiveWorlds. The goal of the users, here is to select all the avatar faces. Their choices are validated for accuracy, thus, preventing unauthorized access to malicious computer programs. The research also, made use of user’s feedback after each test completion and the researchers pointed out that in designing CAPTCHA, a good approach is to make it fun and convenient for users to solve them. Experimental results indicate that it can be solved 62% of the time by human users with an average success time of 24 sec and a positive user rating of 90%. It is designed to be secure against computer programs (bots). Using brute force attack the success rate for a bot to solve it is 1/4096.

Almazayd et al. (2011) proposed a multi-modal CAPTCHA that combined picture and text. An image is being rendered on the screen and many text labels drawn over it. A user has to identify the correct name of the underlying image among the set of text labels that are scattered over it, in order to pass a human verification test. One important feature of this approach is that it renders both the image and text labels together this means that the user has to recognize the image as well as identifying the exact name text-label. The implemented system used thousands of images (animals, fruits, furniture etc.) collected from popular search engines like Google and Bing. A large set of images and text labels are stored in the database and whenever a user tries to access the service, an image is fetched along with four text labels for verification. The use of audio facility in this proposed CAPTCHA scheme was recommended for future research to make it usable for visually impaired users as well as implementing the text-labels in different languages.

Chandakale and Sapkal (2010) proposed a CAPTCHA based algorithm for secure online authentication. The main focus of the proposed system was to provide secure authentication by verifying strength of CAPTCHA of solving or breaking them using pattern matching. Preprocessing, segmentation and character recognition modules for EZ Gimpy CAPTCHA were used. Of the 180 CAPTCHA samples collected from websites, about 154 samples were recognizable with 100% accuracy when tested with the developed system. The researchers concluded that breaking of CAPTCHA can improve development of a robust and secure CAPTCHA for secured online authentication. Their method, however, was not tested on CAPTCHA with connected characters.

Tak et al. (2010), conceptualized asynchronous call from the client side to authenticate the server for stopping the phishing attacks using CAPTCHA. Similarly, James and Philip (2012) presented the use of image CAPTCHA for preventing and detecting phishing based on visual cryptography. It involves a secure way to allow the secret sharing of images without any cryptographic computations. The image CAPTCHA is divided into two shares such that one of the shares is kept with the user and the other share is kept in the server. The user’s share and the original image CAPTCHA is sent to the user for later verification during login phase. After a user logs in by entering his or her confidential information for using account, the user is asked to enter his or her share which is kept with each individual. This share is sent to the server where the user’s share and the share which is stored in the database of the website for each user is stacked together to produce the image CAPTCHA. The end user is required to enter the text displayed in the image CAPTCHA and this can serve the purpose of password and using this, the user can log in into the appropriate website. Using the image CAPTCHA generated by stacking the two shares, one can verify whether the website is genuine and secure or a phishing website and it can also verify whether the user is a human user or not.

Banday and Shah (2009) on clickable image-based CAPTCHA technique involved the use of several sub-images. In their proposed technique a composite CAPTCHA image of a reasonable dimension and resolution is shown to the user. The user has to identify positions of all embedded images that appear as normal with no flip applied to them from the shown composite image. The user needs to click on every non-flipped embedded image to prove human interaction. With this approach it takes less than a second to generate the CAPTCHA image of 240 by 180 pixels size. Accuracy of 96.5 and 93.7% has been obtained for image areas of 240 by 180 pixels with two and four non-flipped sub-images, respectively. Accuracy of 98.5 and 97.50% has been obtained for image size of 480 by 360 pixels with two and four non-flipped sub-images, respectively. User response time for CAPTCHA image of 240 by 180 pixels size is 9.5 sec having 2 non-flipped image and 15.5 sec for image having 4 non-flipped sub-images. This response time decreases to 7.3 and 10 sec for a CAPTCHA image of 480 by 360 pixels. Response time can be further decreased by reducing the degree of distortion.
Sauer et al. (2008) proposed a new CAPTCHA solution that combines pictures of familiar objects with sounds associated with these objects. The goal of their study was to provide insight on the accessibility and usability of audio CAPTCHAs for vision impaired users and to also, inform design of a more robust and accessible CAPTCHA system. Audio CAPTCHAs provide an alternative interface that is accessible to blind users and users with low vision who are unable to see visual CAPTCHAs. A webpage based on the ReCAPTCHA product developed by Carnegie Mellon University was implemented that both a visual distorted text CAPTCHA and an audio CAPTCHA. The proposed expanded prototype increases the number of image/sound combinations to 30 and allowed for multiple sound/image combinations. The researchers reported that the audio CAPTCHA fell well short of the 90% success rate that human users are supposed to have for solving a CAPTCHA. The study showed that a visually impaired user could only solve the CAPTCHA at a rate of 46%, 44% less than what is supposed to be. Also, the average amount of time taken to correctly solve an audio CAPTCHA of 65.64 sec is greater than the 51 sec that is suggested as the time to complete a CAPTCHA. Although, blind user’s times is not expected to rival sighted user times but they could be slight closer.

Fischer and Herfet (2006) presented a simple challenge-response protocol method and prototype for hardening animated visual CAPTCHA, to replace the complex digital signatures, for authentication and security of digital documents. The proposed visual CAPTCHA researchers by obscuring visual documents, so that, their appearances are only recognizable by humans and not computer programs. The researchers noted that there is a very low chance of the document being reconstructed by the computer and this is to ensure that the trusted module which performs the signing, has received the document unmodified. In addition, their approach would improve visual documents integrity and makes it easier to correctly screen human users from automated computer robot access. The limitation of this research is that, the system was not evaluated to determine CAPTCHA complexity, ratio of humans capable of solving the CAPTCHA and the security of this method.

Dailey and Namprendre (2004) proposed a method called Text-Graphics Character (TGC) Completely Automated Public Turing tests to tell Computers and Humans Apart (CAPTCHA) for securing password authenticated systems against dictionary attacks. The aim of this study was to provide a secure login for remote access via consoles or dumb terminal programs that mimics CAPTCHA assisted password authentication despite the difficult challenge of rendering distorted TGC on a text-only screen. A prototype TGC CAPTCHA password authentication method compatible with the SSH user authentication protocol was implemented. In a TGC CAPTCHA-enabled authentication session, the user is challenged by a CAPTCHA test and asked to enter his or her password. If the user passes both the CAPTCHA test and enters the correct password, access is granted. Otherwise, access is denied. But the system was not tested against any kind of attack.

Text based CAPTCHAs have been broken successfully with the help of optical character recognition and every approach is vulnerable against laundry attacks or redirection. The suggested approach ‘Drag and Drop CAPTCHA’ or ‘DnD’ by Tak et al. (2010) is inclusive solution against OCR and laundry attacks. It uses conventional mouse events to recognize human intervention proof. This technique is different in that if bot knows the answer of an artificial intelligence it can’t pass this test without human intervention. It uses familiar action of dragging and dropping items into specific region. In this test user has to solve normal CAPTCHA image but user cannot type the answer of test in to text box. Instead, user has to just drag and drop character blocks in to their respective blank blocks as they appear in the image. Experiments was carried out with approximately 100 people including students, developers and tech-savvies gave good insights of user perspective. The success ratio of the test module is about 87% on first attempt and user satisfaction level was also, high based on feedbacks.

Baird and Bentley (2005) proposed the use of implicit CAPTCHA to distinguish people and machines based on the principle that such challenges are disguised as necessary browsing links challenges can be answered with a single click while still providing several bits of confidence challenges can be answered only through experience of the context of the particular website and challenges are so easy that failure indicates a failed robot attack. Their study summarized the principles underlying implicit CAPTCHAs and illustrated those principles with several images. The researchers also constructed implicit CAPTCHAs that can extract many more bits of confidence by constructing a “story” contained in a sequence of related images (Baird and Bentley, 2005). The user is given an instruction in each image to click on a given subregion for about 8 bits of confidence, a sequence of five such images can then give a total of 40 bits which compares favorably with current explicit CAPTCHAs.
MATERIALS AND METHODS

In this study, we describe the proposed image CAPTCHA authentication system and how it will be applied to successful reduce bot’s access. The design is composed of the implicit CAPTCHA construction and algorithm. By implicit, we mean that the HIP challenge requires human intuition and perception to deduce the expressed challenge. The implicit CAPTCHAs will be woven into the expected sequence of browsing using cues tailored to the site and designed, so that, certain failure modes are correlated with failed bot attacks. This approach is language independent and can be used by wide range of audience irrespective of their familiarity with language and spellings.

Conceptual presentation: Figure 3 depicts the relationships between the different components of the system. It comprises of human users and web bot on the internet trying to get access to secure resources. The image CAPTCHA application server gets images from the local image database which contains images of basic shapes from the internet. The application server then loads these images on the webpage in form of CAPTCHA and requires the users to pass this test to be able to get access to the secure resources they need from the web application.

Implicit CAPTCHA construction: The use of bits of confidence against a random clicker and the retained bits of confidence against an attacker is demonstrated in this section. Supposing a shape to be dragged appears on a 16x4 screen grid as shown in Fig. 4. The black circular ball can be found with a probability of 1/4 on the Horizontal implemented in web applications (H) axis and 1/16 on Vertical (V) the axis. The overall probability of a successful attack on this shape will be computed as 1/64 (1/2^5). This implicit test gives us 6 bits of confidence that such a click is not from a lucky random attacker. Also, there are two potential jump phrases “Welcome” and “Drag circle here” with a probability of 1/2 we retain (2) 1 bit of confidence against an attacker 1 with both OCR and substantial semantic analysis. A shown in Table 1, the implicit image CAPTCHA security increases as the number of jump phrases increases as well as the bit of confidence which implies lower probability of successfully solving of the HIP test by a bot. Multiple image clusters will be used for the CAPTCHA’s HIP test, therefore, the system can retain more bits of confidence against an attacker or bot that can group images. The implication of this is that the likelihood of an automated program taking advantage of the CAPTCHA system is reduced if this drag and drop image CAPTCHA is implemented in web applications.

Fig. 3: Conceptual model
Table 1: CAPTCHA construction

<table>
<thead>
<tr>
<th>Shape coordinate</th>
<th>Bit of confidence from relative distance</th>
<th>No. of jump phrase</th>
<th>Retained bits of confidence for No. of phrases</th>
<th>Total bits of confidence</th>
<th>Probability of bot successful click</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2, 4)</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1/12</td>
</tr>
<tr>
<td>(4, 8)</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>1/32</td>
</tr>
<tr>
<td>(8, 16)</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>10</td>
<td>1/128</td>
</tr>
<tr>
<td>(16, 32)</td>
<td>9</td>
<td>16</td>
<td>4</td>
<td>13</td>
<td>1/512</td>
</tr>
<tr>
<td>(32, 64)</td>
<td>11</td>
<td>32</td>
<td>5</td>
<td>16</td>
<td>1/2048</td>
</tr>
<tr>
<td>(64, 128)</td>
<td>13</td>
<td>64</td>
<td>6</td>
<td>19</td>
<td>1/8192</td>
</tr>
<tr>
<td>(128, 256)</td>
<td>15</td>
<td>128</td>
<td>7</td>
<td>22</td>
<td>1/32768</td>
</tr>
</tbody>
</table>

Fig. 4: Proposed drag and drop CAPTCHA on a 16×4 pixels display

Datasets: This research considered images of shapes, complex backgrounds and varying illuminations. These images were converted to grayscale to help to avoid color-based image recognition algorithms from detecting the unusually bright and uncommon colored shapes and consequently breaking the CAPTCHA.

RESULTS AND DISCUSSION

Performance evaluation: As this is a research in progress, online survey feedback will be used to obtain responses, which will then be analyzed and discussed. The feedback survey will capture some of the following responses from the users, gender, age, education background, previous experience in solving text and image CAPTCHAs, rating the fun factor in solving this drag and drop image CAPTCHA, justifying the choice of shapes, how challenging is it, their preferences in solving text or image CAPTCHAs and usage of this CAPTCHA on their websites. Some of the following parameters will be captured for each participant: success or failure outcome, number of shapes not selected, number of shapes selected, the probability of success, the bit of confidence for the CAPTCHA and the time taken to give the test (Fig. 5).

CONCLUSION

The completed research is expected to contribute to security and authentication in online systems. In an attempt to develop a CAPTCHA system that will be simple but difficult for bots and widely acceptable even among users with low vision, an implicit drag and drop
image CAPTCHA is proposed. This will reduce automated guessing attacks through the use of ‘implicit’ CAPTCHA that requires general human experience but difficult for computer to solve. Rather than relying solely on randomly generated CAPTCHA, random relative positions of the images to the solution will further hardened the CAPTCHA also, by implicit, we mean that the CAPTCHA challenge will always require human intuition and perception as solutions will not be directly expressed in the HIP. By manipulating CAPTCHA coordinates and the number of jump phrases, it is very easy to confuse a bot. Also, solving will involve mouse action and real user interaction that should be challenging for automated bots to mimic as well. This approach will be very useful in processing user’s requests, without any sophistication, in online banking, electronic voting, electronic commerce, health industry among others (Omotosho et al., 2017; Omotosho et al., 2019; Adeyiga et al., 2011; Omotosho et al., 2014; Omotosho et al., 2017).

REFERENCES


