ISSN:0975-9646

Revati Ghadge et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2800-2805

Are you a Human or Robot? or Everything CAPTCHA

Revati Ghadge[#1], Archana M. Naware[#2]

[#1]*Master of Computer Applications (M.C.A.),*
Lokmanya Tilak College of Engineering, Sector 4, Koparkhairane, Navi Mumbai-400709, India
[#2]*Department of Computer Engineering,*
Lokmanya Tilak College of Engineering, Sector 4, Koparkhairane, Navi Mumbai-400709, India

***Abstract*— CAPTCHA stands for Completely Automated Public Turing test to tell Computer and Humans Apart. 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. Bot is a simple computer program used to perform highly repetitive operations. Anti-CAPTCHA technology is advancing, and so are attempts made to build a better CAPTCHA or its equivalent. As the current CAPTCHA methods are striving to be more difficult for bots, they are gradually becoming difficult and annoying for human users as well, and this is the main drawback of CAPTCHA. The applications that use CAPTCHA need to be aware and up to date about reliability of the CAPTCHA they are using. A CAPTCHA system is considered to be broken if an automated attacker reaches a precision of 1%, as it fails the main purpose. To ensure security of the web based applications, users have to prove if they are human or not.**

***Keywords*— CAPTCHA, Turing Test, BOT, Usability, Security, CAPTCHA Alternatives.**

I. Introduction

We often use your computer for website registration, email services, online ticket booking or online voting. However, before the users can move forward with our task, they first have to pass a test. It's not a hard test -- and that's the point. For humans, the test should be simple and straightforward. But for a bot, the test should be almost impossible to solve.

This sort of test is a **CAPTCHA**, also known as a type of **Human Interaction Proof** (**HIP**). Users come across CAPTCHA tests on lots of Web sites. The most common form of CAPTCHA is an image of several distorted letters.

CAPTCHA was invented in 2000 at Carnegie Mellon University (CMU) by Luis Von Ahn, Manuel Blum, Nicholas J. Hooper and John Langford.

CAPTCHA technology has its foundation in an experiment called the **Turing Test**. Alan Turing, known as father of modern computing, proposed the test as a way to examine whether or not machines can think or appear to think like humans. The classic test is a game of imitation, having two participants. One of the participants is a machine and the other is a human. The interrogator can't see or hear the participants and has no way of knowing which is which. If the interrogator is unable to figure out which participant is a machine based on the responses, the machine passes the Turing Test [1].

The goal of **CAPTCHA**, is to create a test that humans can pass easily but machines cannot. It is also important that the CAPTCHA application is able to present different CAPTCHAs to different users. If a visual CAPTCHA presented is same for every user, a spammer will spot the form, decipher the letters and will program an application to type in the correct answer automatically.

Who Uses CAPTCHA? CAPTCHAs are mainly used by websites that offer services like online polls and registration forms. Web-based email services like Gmail, Yahoo and Hotmail offer free email accounts for their users. On each sign-up process, users will come across a CAPTCHA at the end of the sign-up form so as to ensure that the form is filled out only by a legitimate human and not by a computer bot. CAPTCHAs are used to prevent spammers from using a bot to generate hundreds of spam mail accounts [2].

II. Types of CAPTCHA

CAPTHCAs mean presenting a challenge response test to the users. Most commonly used, following are the types of captcha:

1. The Standard Distorted Text CAPTCHA with an Audio Option

Figure 1. Distorted Text CAPTCHA

It is reliable, but some of the distorted word images are rather hard to solve. To get past that it allows you the option to “reCaptcha”, in order to receive a new one. There is also an audio option if you are unable to visually make out the word.

2. Picture Identification Captcha

Figure 1. Picatcha

This CAPTCHA provides the user with a choice of choosing the correct image that they are asked to identify.

They never get harder than basic images so users do not have worry about not being able to depict the difference between them and the incorrect images.

3. Math Solving Captcha

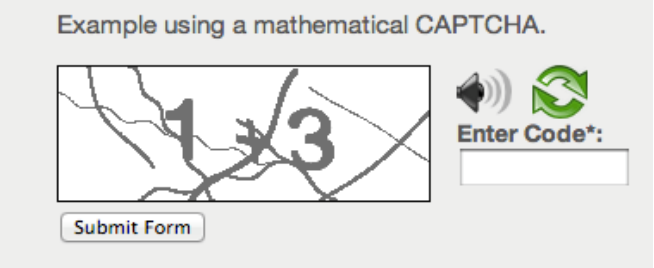

Figure 2. Math Solving Captcha

Users have to solve basic math problems to carry on with their tasks in this type of CAPTCHA.

4. NO CAPTCHA reCAPTCHA

What is reCAPTCHA? reCAPTCHA is a free service to protect your website from spam and abuse. reCAPTCHA uses an advanced risk analysis engine and adaptive CAPTCHAs to keep automated software from engaging in abusive activities on your site. It does this while letting your valid users pass through with ease [3].

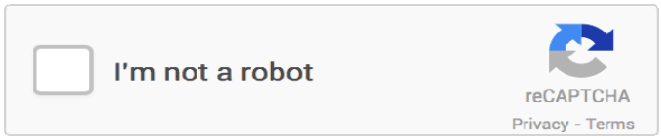

Figure 3.1. NO Captcha

This is the most recent development for bot protection called “NO CAPTCHA”. Here, they directly ask the users whether they are robots or not? And they user just has to do nothing more than select a single checkbox next to the statement “I’m not a robot” [4].

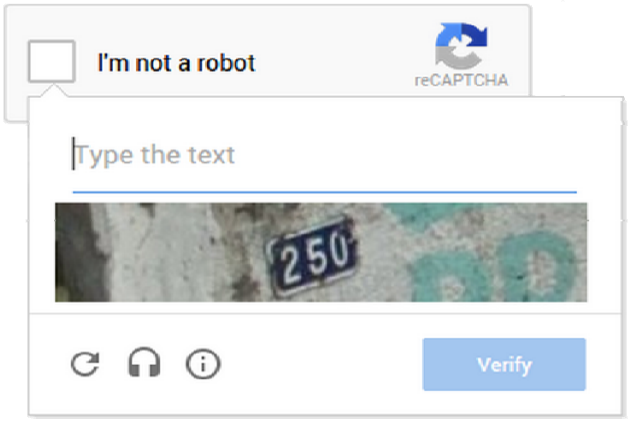

Figure 4.2. reCaptcha with Text Captcha

However, CAPTCHAs aren't going away just yet. In cases when the risk analysis engine can't confidently predict whether a user is a human or an abusive agent, it will prompt a CAPTCHA to elicit more cues, increasing the number of security checkpoints to confirm the user is valid.

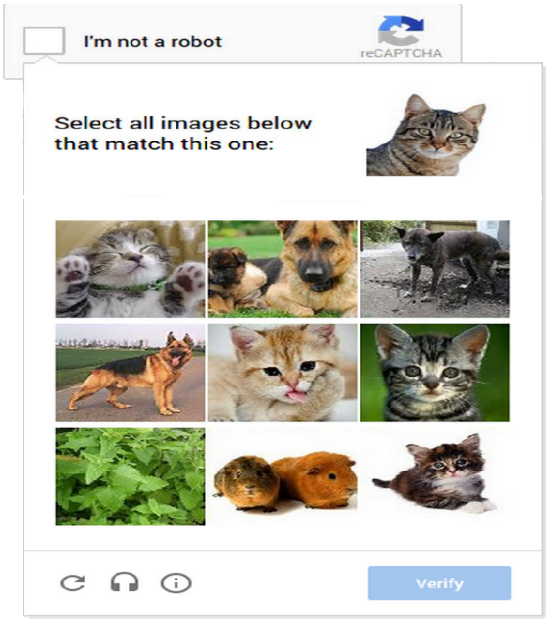

Figure 4.3. reCaptcha with Image Captcha

5. Social Authentication Captcha

Figure 4. Facebook Friend Recognition Captcha

Instead of showing you a traditional captcha on Facebook, one of the ways they verify your identity is through social authentication. It shows you a few pictures of your friends and ask you to name the person in those photos. Hackers halfway across the world might know your password, but they don't know who your friends are. This might create problems for those who has long list of social friends and they even don’t know the name of all or are not able to remember them [5].

III. APPLICATIONS OF CAPTCHA

1. Online Polls: Bots can wreak havoc to any unprotected online poll. They might create a large number of votes which would then falsely represent the poll winner in spotlight. This also results in decreased faith in these polls. CAPTCHAs can be used in websites that have embedded polls to protect them from being accessed by bots, and hence bring up the reliability of the polls [6].

2. Protecting Web Registration: Several companies offer free email and other services. Bots can take advantage of the service and can sign up for a large number of accounts. This often creates problems in account management and also increased the burden on their servers. CAPTCHAs can effectively be used to filter out the bots and ensure that only human users are allowed to create accounts [6].

3. Preventing comment spam: Most bloggers are familiar with programs that submit large number of automated posts that are done with the intention of increasing the search engine ranks of that site. CAPTCHAs can be used before a post is submitted to ensure that only human users can create posts [6].

4. E-Ticketing: Ticket brokers also use CAPTCHA applications. These applications help prevent ticket scalpers from bombarding the service with massive ticket purchases for big events. Legitimate customers become victims as events sell out minutes after tickets become available. Scalpers then try to sell the tickets above face value [6].

5. Email spam: CAPTCHAs also present a plausible solution to the problem of spam emails. All we have to do is to use a CAPTCHA challenge to verify that an indeed a human has sent the email [6].

What CAPTCHA Does Not Do?

It can't keep humans with spam intentions from mucking things up. A person can still bypass the CAPTCHA system and perform spam like operations that the websites have to moderate later.

IV. USABILITY ISSUES

It is widely accepted that a good CAPTCHA must be both robust and usable. The robustness of a CAPTCHA is its strength in resisting adversarial attacks. It is observed that overall usability of CAPTCHA decreases with increase in complexity.

Usability of CAPTCHA is determined by the following factors:

1. Accuracy: Accuracy with which user can pass a CAPTCHA challenge. For example, the number of times user tries to pass the CAPTCHA
2. Response time: Time taken by a user to pass the test
3. Perceived difficulty/satisfaction of using a scheme: How difficult to use do people perceive a CAPTCHA is? Are users subjectively satisfied and would they be willing to use such a scheme? [7]

Usability issues commonly affecting the users considering the above factors are:

1. **Distortion:**

   Distortion has a clear impact on the usability of CAPTCHAs, since human users would find it difficult or impossible to recognize over-distorted characters. To cope with usability problems caused by distortion, a system will have to allow multiple attempts for each user. Typically a new challenge is used for each attempt. This will not only annoy users, but also lowers the security of the system by a factor of the number of allowed attempts. Distortion often creates ambiguous characters, where users cannot be sure what they are [7].

   One of the major problems with distorted captchas is that they are very difficult or impossible for people with reduced vision.

   In audio CAPTCHAs, letters are read aloud instead of being displayed in an image. Typically, noises are deliberately added to prevent such audio schemes from being broken. Background noises effectively distort sounds in audio CAPTCHAs.

Figure 5. Example of Distortion

2. **Content:**

   The choice of content materials used in each CAPTCHA challenge can also have significant impact on usability. Depending on the font the captcha uses, a lower-case "l" as in "lama" can look exactly the same as an uppercase "I" as in Iguana or the number "1".It can also be very difficult to tell the difference between an uppercase letter "O" as in "Ocean" and the numeral "0" or zero.

   Content materials used in both text and audio CAPTCHAs are typically language specific. Digits and letters read in a language are often not understandable to people who do not speak the language. Localization is a major issue that CAPTCHAs face [7].

3. **Presentation:**

   Color is extensively used in user interfaces. Color schemes might also be expected to work as an additional defense against software attacks in some schemes, since typically bots performs poorly in recognizing texts in color [7].

   The audio CAPTCHAs can cause compatibility issues with web pages as they might require additional plug-ins in order to be used.

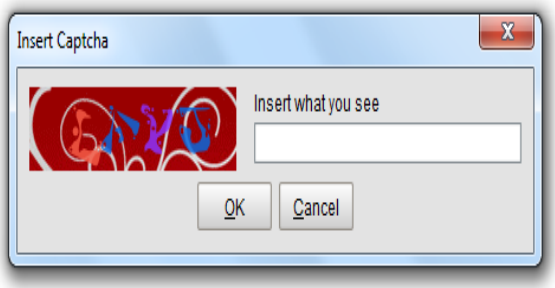

Figure 6. Example of Presentation

The design of CAPTCHA is an art, rather than a science. It requires considerable study to evolve the design of secure and user friendly CAPTCHAs.

V. CHALLENGES FOR CAPTCHA

The challenge in breaking a CAPTCHA is teaching a computer how to process information in a way similar to how humans think.

Californian startup company, Vicarious claims to have developed the first artificial intelligence software capable of reliably solving most modern CAPTCHAs. The Vicarious AI achieves success rates of up to 90% on a wide range of the most common CAPTCHAs, including those from Google, Yahoo, PayPal, Captcha.com, and others [8].

If CAPTCHA-breaking technology like the one developed by Vicarious became widely available, in the future the results would be disastrous for the hundreds of thousands of websites that rely on a CAPTCHA system to fend off content spammers. Sites with user engagement components would be so overrun with spam comments and fake registrations that they would likely have to resort to shutting off user engagement altogether, unless they came up with an alternate security system [9].

A smart move is to consider a CAPTCHA system to be broken if an automated attacker can reach a precision of 1%. Suppose, the CAPTCHA breaking reaches 90% precision, it means that CAPTCHA has been broken pretty severely.

The reliability of these algorithms is not assured though. However, Spammers can afford to have only one-third of their attempts succeed if they set bots to break CAPTCHAs several hundred times every minute.

Luis Von Ahn one of the inventors of CAPTCHA talked about the relationship between things like CAPTCHA and the field of artificial intelligence (AI) in a lecture at CMU in 2006. CAPTCHA is a barrier between spammers or hackers and their goal, these people have dedicated time and energy toward breaking CAPTCHAs. Their successes mean that machines are getting more sophisticated. Every time someone figures out how to teach a machine to defeat a CAPTCHA, we move one step closer to artificial intelligence [2].

As computers become more sophisticated, the testing method must also evolve. But if the test evolves to the point where humans can no longer solve a CAPTCHA with a decent success rate, the system as a whole fails. The answer may not involve warping or distorting text -- it might require users to solve a mathematical equation or answer questions about a short story. And as these tests get more complicated, there's a risk of losing user interest. Not many people will want to post a reply to a message board if they must first solve a quadratic equation. [2]

Eventually, we might reach a point where computers and humans perceive puzzles the same way. If that happens, tests like CAPTCHA will become useless lines of code.

VI. ALTERNATIVES

Anti-CAPTCHA technology is advancing, but so are attempts to build a better “are you really a human?” tests.

The PlayThru authentication module from **Are You a Human** displays a very, very simple game, which is different each time. For example, it might display a number of floating objects and ask you to put only the tools in the toolbox, or put toppings on a pizza. Win the game and you’ve authenticated yourself as a human [10].

Figure 8. Are you a human?

Honeypots are traps made to catch bots without ever being noticed by human users. The most common example is the hidden form field. With this solution, an extra field is included in the web form and then hidden from human users with JavaScript or CSS. Bots, however, will still “see” the field and fill it out. If the field is filled out, the form is automatically rejected [11].

One other CAPTCHA alternative is the **Timing Trick**. As humans, it usually takes us a few seconds or minutes to fill out most Web forms, but bots complete them instantaneously. Some security experts are using this trait against bots, setting a minimum amount of time to complete forms. Any bot trying to instantaneously submit a form is rejected as obviously not human [12].

Minteye’s offering combines a CAPTCHA replacement with built-in advertising. It displays an image that’s been distorted by swirling it around the center, along with a slider that adjusts the degree of swirl. When you click the button with the slider at the zero-spot where the distortion is gone, you’ve solved it [10].

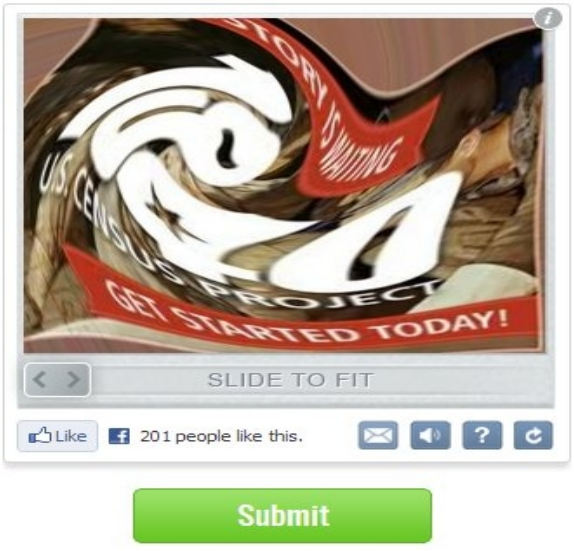

Figure 9. Swirl

However, there have been multiple reports for this one on the internet stating as to how easily it can be hacked.

A team at Carnegie Mellon is working on a new test called a GOTCHA test, which asks users to describe patterns seen in inkblots. But it's not ready for primetime yet [9].

This new approach relies on a user answering a number of questions when he or she first signs up for access to a website. It begins by generating a set of simple inkblot pictures by randomly positioning different coloured ink spots in a small area of the screen (Figure 10).

This inkblot test is based on the Rorschach test, which is a method of psychological evaluation. Psychologists use this test in an attempt to examine the personality characteristics and emotional functioning of their patients [13].

This new test is called **GOTCHA** which stands for Generating panOptic Turing Tests to Tell Computers and Humans Apart.

The human ability to recognise patterns far outstrips anything that computers can do and matching this with the user's interpretation of random patterns is clever. There is plenty of evidence that pattern recognition is easier than remembering passwords [14].

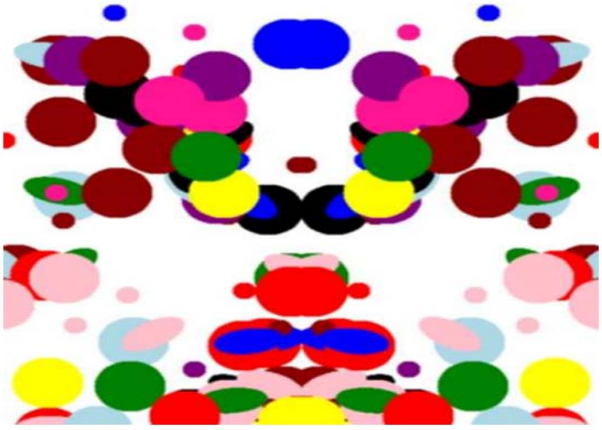

Figure 10. Inkblot

In Figure 11, when the user registers for the first time, he will set the phrases against each GOTCHA (inkblot). When user clicks on the Generate Password File button a password file is generated to store all the user responses which is used to authenticate the user whenever the user tries to login [15].

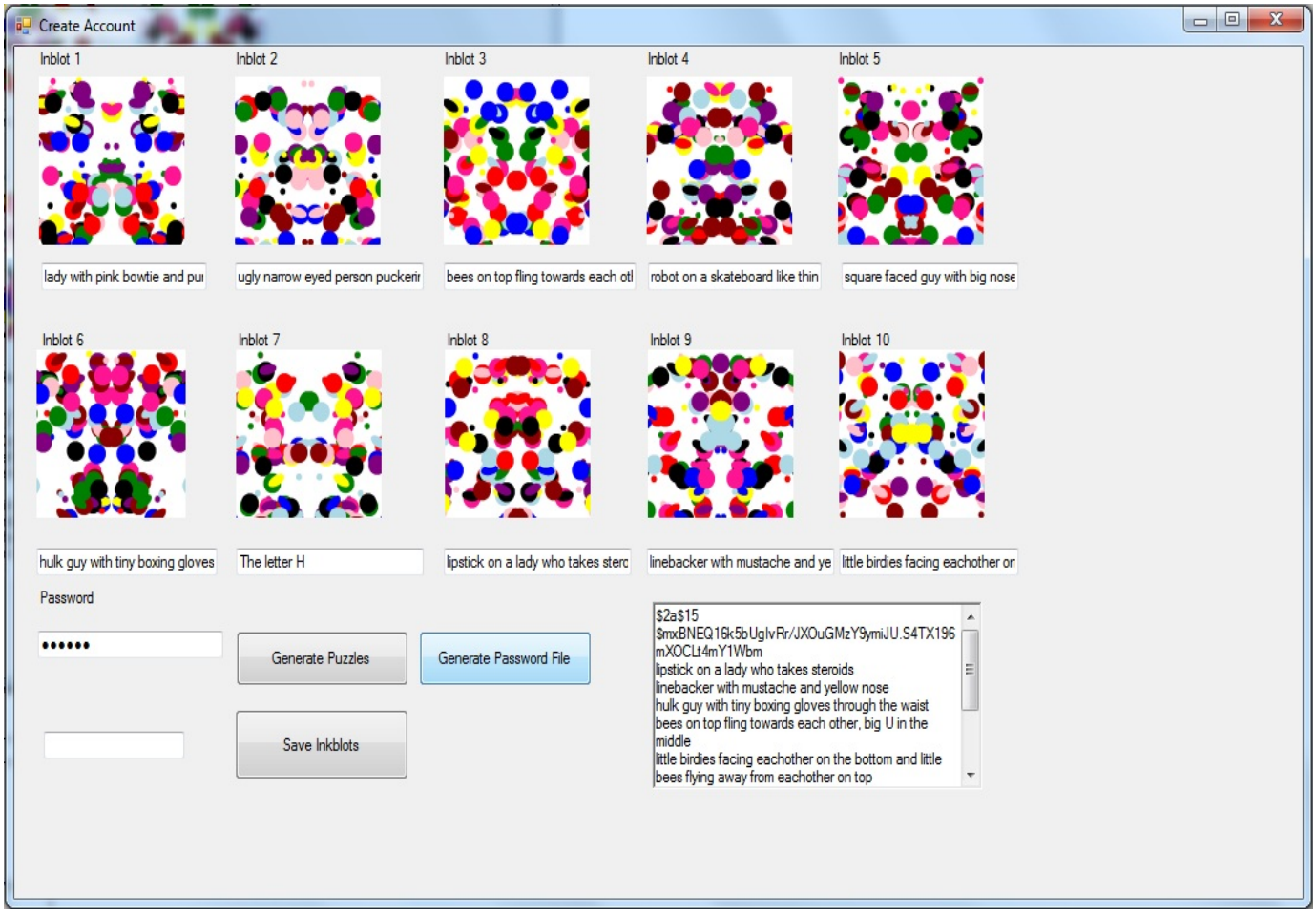

Figure 11. Set phrases against the inkblots

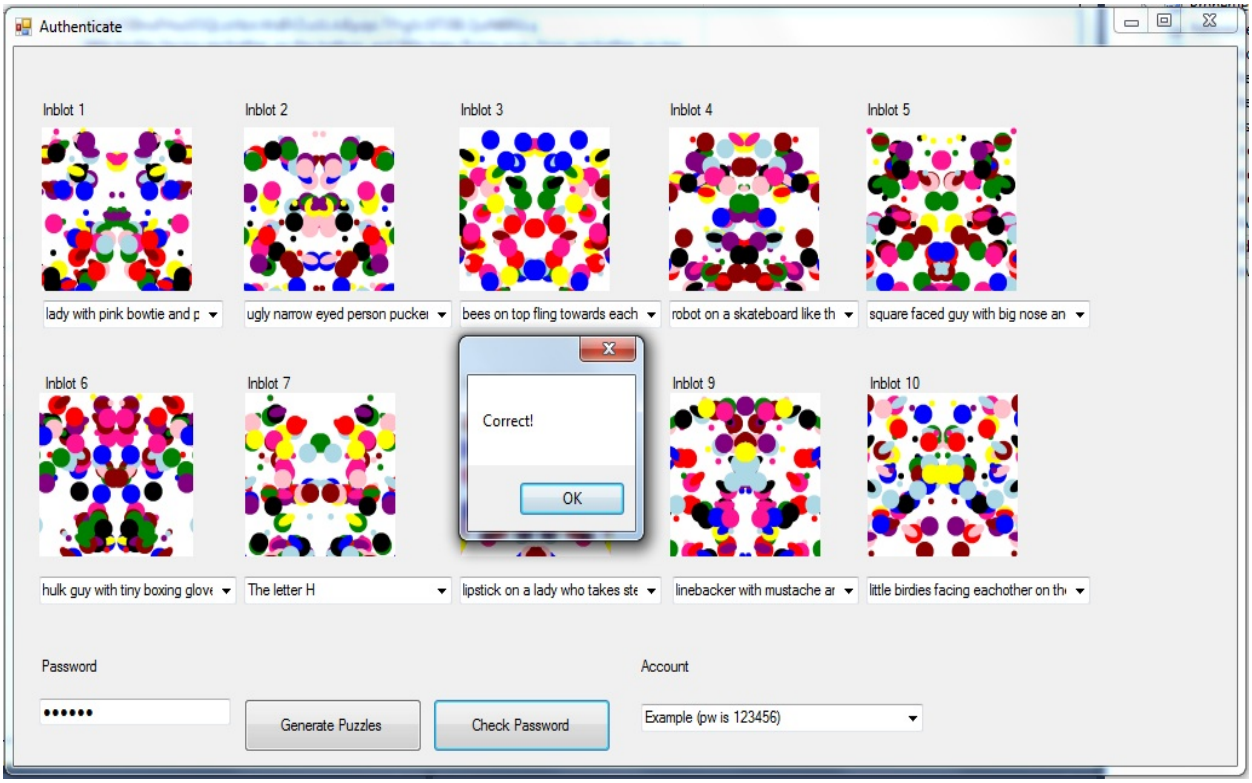

Figure 12. Authenticate on each login

In Figure 12, we see the authentication screen. To authenticate the user would use the form Authenticate. When the user logs in, he will enter the password, and click Generate Puzzles. Finally, the user would match the labels with the appropriate Inkblots and click Check Password [15].

But it does raise the question that how well people will remember their original interpretation of an inkblot. This may well be highly dependent on their state of mind at the time which in turn which could be influenced by all kinds of local and temporary variables [14].

The web security can be increased by an additional security level in the web sites before the CAPTCHA, such as Honeypots or Timing Trick. Many other developers are working to solve the problem of separating humans from bots without annoying the humans.

VII. Conclusion

CAPTCHA has become a necessity for security of various websites and its users. Although humans may find it annoying, CAPTCHAs help reduce bot attacks.

This paper tries to focus on the fact that even though most of the human population considers CAPTCHA a major inconvenience, it is important for protection from bots. Looking at the scenario on the web, it is safe to say that CAPTCHAs are not going away any time soon.

In a perfect future world, we'd each boast a unique electronic identity like the security number that would be impossible to forge and accepted by every application and website [13]. Till then, we identify ourselves with passwords (strong ones, of course) and by solving CAPTCHAs or equivalents to continue proving we are human!

References

[1] "Turing Test", http://en.wikipedia.org/wiki/Turing_test
[2] "How Captcha works", http://computer.howstuffworks.com/captcha.htm
[3] "reCaptcha", http://www.google.com/recaptcha/
[4] "NO CAPTCHA RECAPTCHA", http://googleonlinesecurity.blogspot.in/2014/12/are-you-robot-introducing-no-captcha.html
[5] "Need of Captcha", http://www.logicmatters.org/news/its-all-about-captcha/
[6] "CAPTCHA", http://captcha.net/
[7] Jeff Yan, Ahmad Salah El Ahmad "Usability of CAPTCHAs"
[8] "Breaking the Captcha", http://www.itproportal.com/2013/10/28/ai-startup-vicarious-claims-breakthrough-following-captcha-test-crack/
[9] "Future of Captcha", https://www.mollom.com/blog/the-future-of-captchas
[10] "Alternatives for Captcha", http://www.itproportal.com/2013/02/15/are-you-a-human-captcha-and-the-future/
[11] http://www.usertesting.com/blog/2014/04/09/think-your-site-needs-captcha-try-these-user-friendly-alternatives/
[12] http://www.scientificamerican.com/article/pogue-8-alternatives-to-hated-captcha/
[13] http://theinkblot.com/
[14] "GOTCHAs and the future of Captchas", http://www.technologyreview.com/view/520306/will-gotchas-replace-captchas/
[15] "GOTCHA Challenge", http://www.cs.cmu.edu/~jblocki/GOTCHA-Challenge.html