



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

reCAPTCHA and its Implementation: Human-Based Character Recognition via Web Security Measures

Neha Vij¹

Student, Masters in Computer Science, California State University Sacramento, California, USA¹

ABSTRACT: reCAPTCHA is security measure used to protect websites against automated abuse by presenting random auto-generated challenge for users to solve. These challenges are designed in such a way it is difficult for computers to solve yet easy for humans. In this paper, we presented brief introduction to CAPTCHA, types of CAPTCHA, advantages, disadvantages of CAPTCHA and why we chose reCAPTCHA with valid reasons. Explained reCAPTCHA with challenge types and then implemented how to create a reCAPTCHA for any website. We also presented advantages of reCAPTCHA in security, usage, value of the current trends.

KEYWORDS: CAPTCHA, reCAPTCHA, Bots.

I. INTRODUCTION

reCAPTCHA is a recent development of CAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart) that was proposed in 2008 by Luis von Ahn to invest the numerous human computational power in solving problems that computers cannot solve yet. A CAPTCHA is a challenge response test that is used on World Wide Web to determine whether a particular user is human or computer. A CAPTCHA is an image which contains several distorted characters that usually appears at the bottom of web registration forms. Users are asked to type such distorted characters to prove themselves as human user which is typically not possible by computer, so this will prevent abusing online services. CAPTCHAS are used by many websites and blogs. CAPTCHAS were used by too many users per day. Though CAPTCHAS prevent lot of online services abuse the mental effort spent by each person is wasted due to complexity sometimes. CAPTCHAS provide same level of difficulty for all kind of users which is waste of time to do such computation for trusted users. So then google introduced recent improvement on CAPTCHA i.e. reCAPTCHA. The new reCAPTCHA allows number of users to verify themselves without solving a CAPTCHA. This is done by just single click confirming that they are not robot. This is being called NO CAPTCHA reCAPTCHA experience. ReCAPTCHA is free service that protects too many websites from spam and abuse.

II. MORE ON CAPTCHA

A CAPTCHA is a program that will protect websites against bots by generating tests that only humans can pass but current computer programs cannot. For example, humans can read distorted texts, but computer programs cannot. CAPTCHAS was first put to use in early 2000's by the reputed search engine giants Alta Vista and Yahoo.

Types of CAPTCHA:

1. Text-based Captcha: This is the most common type of captcha used where text is in distorted form.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

Type the characters you see in the picture below.


Letters are not case-sensitive

Fig 1: Text Captcha

2. Recaptcha: It is mix of distorted text and audio. If there's any possibility that we are not able to solve through distorted text then audio option can be used to solve the captcha.



Fig 2: Recaptcha

3. Mathematical Captcha: Presented with some mathematical problem to the user to solve.

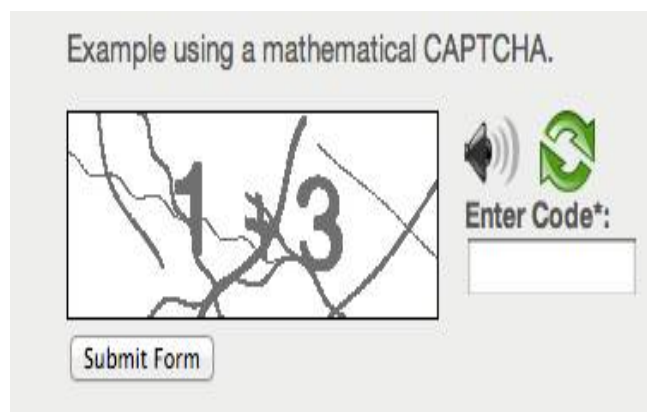


Fig 3: Mathematical Captcha

4. Image Based Captcha: In image based captcha several images are juxtaposed and actions like click the images of particular item is given to the user.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

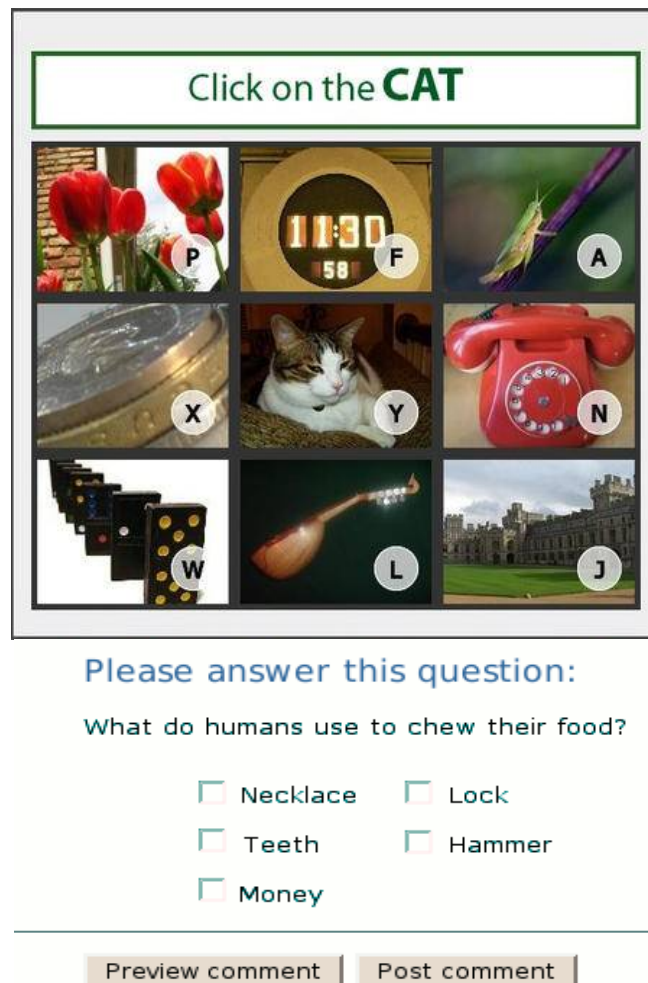


Fig 4: Logic Based captcha

5. Commercial Captcha: These are also called as ad-injected Captcha. Commercial captchas in-fact help website owners earn some extra money every time the ad is clicked. The distorted text should be deciphered and entered correctly in the space provided and more often than not, the space to enter the text is provided just beneath the displayed ad. There are two benefits- You make some decent revenue and at the same time, brand promotion and brand recognition become all the more impactful. Of-course, web spam attacks are also obviated.



Fig 5: Ad-Injected Captcha

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 5, Issue 11, November 2017

6. 3D Captcha: 3D Captchas are eye-straining and nerve-wracking. Leave apart automated chat bots and other malicious programs, even humans find 3D Captchas difficult. 3D Captchas are also called 'Super Captchas' for their 3D images and words.



Fig 6: 3D Captcha

Advantages of Captcha:

1. Preventing Comment Spam in Blogs: Most bloggers are familiar with programs that submit bogus comments, usually for the purpose of raising search engine ranks of some website (e.g., "buy penny stocks here"). This is called comment spam. By using a CAPTCHA, only humans can enter comments on a blog. There is no need to make users sign up before they enter a comment, and no legitimate comments are ever lost!
2. Protecting Website Registration: Several companies (Yahoo!, Microsoft, etc.) offer free email services. Up until a few years ago, most of these services suffered from a specific type of attack: "bots" that would sign up for thousands of email accounts every minute. The solution to this problem was to use CAPTCHAs to ensure that only humans obtain free accounts. In general, free services should be protected with a CAPTCHA in order to prevent abuse by automated scripts.
3. Protecting Email Addresses From Scrapers: Spammers crawl the Web in search of email addresses posted in clear text. CAPTCHAs provide an effective mechanism to hide your email address from Web scrapers. The idea is to require users to solve a CAPTCHA before showing your email address.
4. Online Polls: In November 1999, <http://www.slashdot.org> released an online poll asking which was the best graduate school in computer science (a dangerous question to ask over the web!). 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. However, students at Carnegie Mellon found a way to stuff the ballots using programs that voted for CMU thousands of times. CMU's score started growing rapidly. The next day, students at MIT wrote their own program and the poll became a contest between voting "bots." MIT finished with 21,156 votes, Carnegie Mellon with 21,032 and every other school with less than 1,000. Can the result of any online poll be trusted? Not unless the poll ensures that only humans can vote.
5. Preventing Dictionary Attacks: CAPTCHAs can also be used to prevent dictionary attacks in password systems. The idea is simple: prevent a computer from being able to iterate through the entire space of passwords by requiring it to solve a CAPTCHA after a certain number of unsuccessful logins. This is better



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 5, Issue 11, November 2017

than the classic approach of locking an account after a sequence of unsuccessful logins, since doing so allows an attacker to lock accounts at will.

6. Search Engine Bots: It is sometimes desirable to keep webpages unindexed to prevent others from finding them easily. There is an html tag to prevent search engine bots from reading web pages. The tag, however, doesn't guarantee that bots won't read a web page; it only serves to say "no bots, please." Search engine bots, since they usually belong to large companies, respect web pages that don't want to allow them in. However, in order to truly guarantee that bots won't enter a web site, CAPTCHAs are needed.
7. Worms and Spam: CAPTCHAs also offer a plausible solution against email worms and spam: "I will only accept an email if I know there is a human behind the other computer." A few companies are already marketing this idea.

Disadvantages of Captcha:

1. Sometimes very difficult to read.
2. Are not compatible with users with disabilities
3. Time-consuming to decipher.
4. Technical difficulties with certain internet browsers.
5. May greatly enhance Artificial Intelligence.

III. RECAPTCHA

reCAPTCHA is a free service that protects 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.

reCAPTCHA offers more than just spam protection. Every time our CAPTCHAs are solved, that human effort helps digitize text, annotate images, and build machine learning datasets. This in turn helps preserve books, improve maps, and solve hard AI problems.

The reCaptcha service offered by Google, is the most widely used captcha service, and has been adopted by many popular websites for preventing automated bots from conducting nefarious activities. Google announced that deployment of a new reCaptcha mechanism designed to be more human-friendly and secure.

Widget

When visiting a webpage protected by reCaptcha, a widget is displayed (shown in Figure 7). The widget's JavaScript code is obfuscated, to prevent analysis from third parties. When the widget loads, it collects information about the user's browser which will be sent back to the server. Furthermore, it performs a series of checks for verifying the user's browser.

Workflow

Once the user clicks in the checkbox, a request is sent to Google containing (i) the Referrer, (ii) the website's site key (obtained when registering for reCaptcha), (iii) the cookie for google.com, and (iv) the information generated by the widget's browser checks (encrypted). The request is then analysed by the advanced risk analysis system, which decides what type of captcha challenge will be presented to the user.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 5, Issue 11, November 2017

Challenge Type

The different type of challenges varies from user to user. Harder challenges will be presented if a specific user has low reputation or requests multiple challenges or provides several wrong answers many times. In our experiments we came across the following versions of reCaptcha:

No captcha reCaptcha: The new user-friendly version is designed to completely remove the difficulty of solving captchas. Upon clicking the checkbox in the widget, if the advanced risk analysis system considers the user have high reputation, the challenge will consider to be solved and no action required from the user.

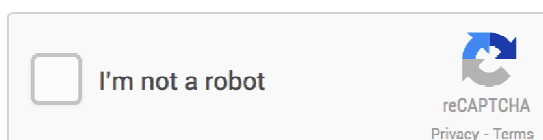


Fig 7: Before user clicks checkbox

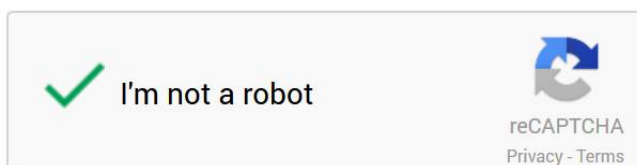


Fig 8: User considered Human

Image reCaptcha: This new version is built on the notion that identifying images with similar content. The challenge contains a sample image and 9 candidate images, and the user is requested to select those that are similar to the sample. The challenge usually contains a keyword describing the content of the images that the user is required to select. The number of correct images varies between 2 and 4.



Fig 9: Image reCaptcha

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 5, Issue 11, November 2017

Text reCaptcha: These distorted texts are returned when the advanced risk analysis consider the user to have a lower reputation. Text reCaptcha is fall back captcha which will be selected when the User-Agent fails certain browser checks, the widget automatically fetches and presents a challenge of this type, before the checkbox is clicked. Over the period of the following 6 months, text captchas appeared to be gradually “phased out”, with the image captcha now being the default type returned, as these captchas are harder for humans to solve despite being solvable by bots.

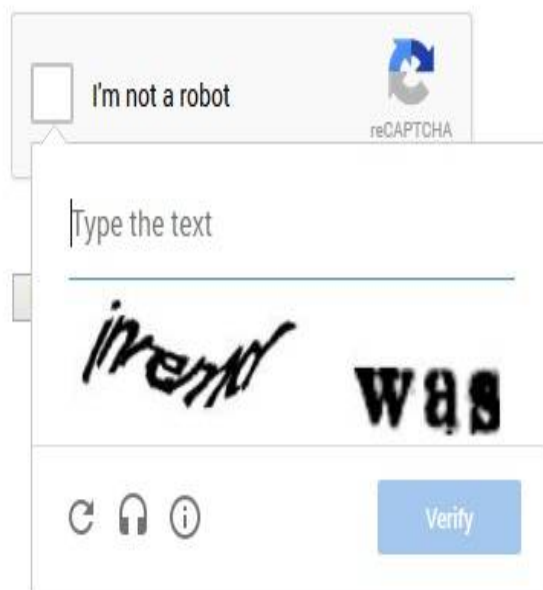


Fig 9: Text reCaptcha

IV.IMPLEMENTATION

To register for a new site we need to fill a form wherein we first choose the type of reCAPTCHA and it can be any from these: reCAPTCHA V2, Invisible reCAPTCHA, reCAPTCHA Android Library. User is required to fill in the list of domains. It can be localhost or machine name. First, choose the type of reCAPTCHA and then fill in authorized domains or package names. After you accept our terms of service, you can click Register button to get new API key pair. The reCAPTCHA site key created will only work on these domains. Following are the steps to add reCAPTCHA to your site or mobile application:

1. Choose the client side integration:
 1. reCAPTCHA V2
 2. Invisible reCAPTCHA
 3. reCAPTCHA Android Library
2. Verifying the user's response

To start using reCAPTCHA, we need to sign up for an API key pair for your site. The key pair consists of a site key and secret key. The keys generated here will be used to verify users in the next steps. The site key is used to invoke reCAPTCHA service on your site or mobile application. The site key provided in this code lab can work only on localhost or any other live website which user wants to hosts. The secret key authorizes communication between your



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 5, Issue 11, November 2017

application backend and the reCAPTCHA server to verify the user's response. The secret key needs to be kept safe for security purposes.

Creating a Login page:

- 1) User creates a page Login.html page containing 2 fields: username and password. User adds site key to the div tag. This site is the one which user received after registering.
- 2) This request goes to the Login Servlet where username and password is hardcoded and from where it receives the users input.
- 3) In web.xml, Login Servlet needs to be mapped.
- 4) It goes to verify method VerifyRecaptcha which is a JSON request and it's validated and sent back as a response to Login Servlet.
- 5) If username and password matches then the request is forwarded to LoginSuccess.jsp which gives success message.

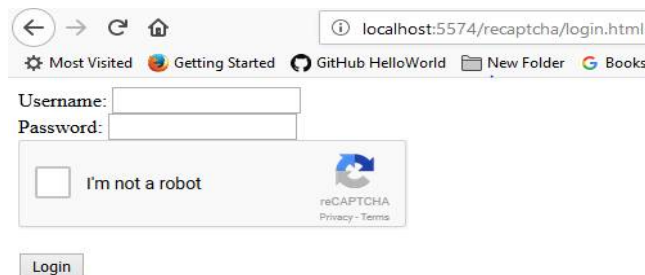


Fig 10: Login Page for user



Fig 11: Captcha not selected

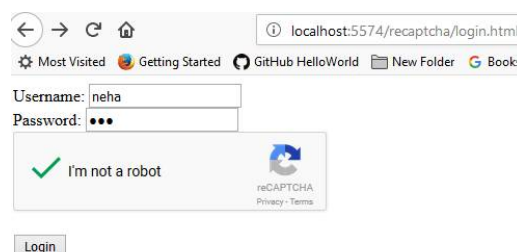


Fig 12: User Validated



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 5, Issue 11, November 2017

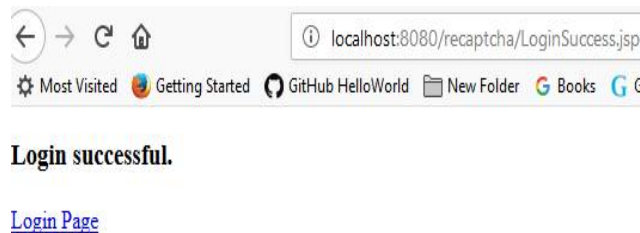


Fig 13: User login successful

V.ADVANTAGES OF RECAPTCHA

a. ADVANCED SECURITY

1. reCAPTCHA protects and defends - reCAPTCHA is built for security. Armed with state of the art technology, it always stays at the forefront of spam and abuse fighting trends. reCAPTCHA is on guard for you, so you can rest easy.
2. Not just distorted text - reCAPTCHA doesn't depend solely on text distortions to separate humans from machines. Rather it uses advanced risk analysis techniques, considering the user's entire engagement with the CAPTCHA, and evaluates a broad range of cues that distinguish humans from bots.
3. Bots Beware - reCAPTCHA is the most widely used CAPTCHA provider in the world. Our broad installed publisher base provides an unparalleled view into abusive activity on the internet, so the bad guys cannot hide. reCAPTCHA knows when to be tough to keep the bots at bay.

b. Ease of Use

1. Easy for People. Hard for Bots - Purposefully designed. And actively aware. reCAPTCHA knows when to be easy on people and hard on bots.
2. Accessibility Option - Safe & Sound - The reCAPTCHA project does make efforts to provide audio alternatives for visually impaired users, but many more text-recognition CAPTCHAs are being used without aids. For the visually impaired, reCAPTCHA provides an audio CAPTCHA option. Your legitimate users will find our regular audio CAPTCHAs easy to solve. Bots on the other hand will get a much harder audio CAPTCHA designed to block them.

c. creation of value

Every time our CAPTCHAs are solved, that human effort helps digitize text, annotate images, and build machine learning datasets. This in turn helps preserve books, improve maps, and solve hard AI problems.

1. One CAPTCHA at a time - Millions of CAPTCHAs are solved by people every day. reCAPTCHA makes positive use of this human effort by channelling the time spent solving CAPTCHAs into digitizing text, annotating images, and building machine learning datasets. This in turn helps preserve books, improve maps, and solve hard AI problems.

2. Stop a bot. Improve a map - reCAPTCHA improves our knowledge of the physical world by creating CAPTCHAs out of text visible on Street View imagery. As people verify the text in these CAPTCHAs, this information is used to make Google Maps more precise and complete. So if you're a Google Maps user, your experience (and everyone else's) will be even better.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 5, Issue 11, November 2017

3. Build a bot - reCAPTCHA helps solve hard problems in Artificial Intelligence. High quality human labelled images are compiled into datasets that can be used to train Machine Learning systems. Research communities benefit from such efforts that help build the next generation of ground breaking Artificial Intelligence solutions.

4. Save a book - reCAPTCHA digitizes books by turning words that cannot be read by computers into CAPTCHAs for people to solve. Word by word, a book is digitized and preserved online for people to find and read. reCAPTCHA uses scanned text that optical character recognition (OCR) technology has failed to interpret. This, in theory, provides unbreakable Captchas with the secondary benefit of helping to digitize books.

VI.CONCLUSION

reCAPTCHA technology was developed not merely with an eye toward improving cyber security, but also as a way to harness and reuse the collective human time and mental energy spent solving and typing Captchas. It needs certain improvement. There is no option to hide it for logged in users (say Administrator) neither it can be customized according to website color and texture. It is not accessible and convenient to use. Standard Recaptcha has been identified as one of the greatest barriers to accessibility on the modern web by accessibility advocates. Standard Recaptcha makes your form difficult enough to submit as it is. Visual captchas take 10 seconds for a user to complete, audio captchas almost 30 seconds. Users who do not speak English as their primary language have to devote 20 seconds to solving an English captcha. Some simple changes could be instituted to make reCAPTCHA both ethical and (potentially) legal. reCAPTCHAs should make explicit the fact that the user is doing work. This would resolve some minor legal issues but ameliorate major ethical issues as seen within several ethical frameworks. Doing this is the first step to making consent possible. The other major change that, in addition to the above, would make reCAPTCHA ethically sound and in better legal standing would be to make it opt-in (e.g., by making it clear to humans doing an OCR task which word is the reCAPTCHA word and which is the CAPTCHA word). But only single registration of domain name is enough for sub-domains or sub-directories which is an advantage. Web sites adopt reCAPTCHA, although reCAPTCHA presents two words instead of just one, it typically takes no more time for users to solve a reCAPTCHA than to solve a standard CAPTCHA. We can conclude by saying that by making some relatively minor adjustments to existing reCAPTCHA systems, human brainpower can be used ethically and legally and it shouldn't be used unless there is any compelling reason to do so.

REFERENCES

- [1] I'm not a human: Breaking the Google reCAPTCHA SuphaneeSivakorn, Jason Polakis, and Angelos D. Keromytis
- [2] L. V. Ahn, B. Maurer, D. Abraham and M. Blum, "reCAPTCHA: Human-based character recognition via web security measures," Science Magazine, pp. 1465 - 1468, 12 September 2008.
- [3] reCAPTCHA google developers: <https://developers.google.com/recaptcha/intro>
- [4] P. Gutmann, D. Naccache and C. C. Palme, "CAPTCHAs: Humans vs. Bots", IEEE Security & Privacy, 2006, pp. 68-70.
- [5] Google Inc., "reCAPTCHA: Stop spam, read books,"
- [6] WordPress "reCAPTCHA Definition history"
- [7] <https://www.tutorialspoint.com/articles/the-story-of-captcha-and-its-variants>
- [8] <http://www.captcha.net/>
- [9] <http://www.google.com/recaptcha/faq>
- [10] JonathanLung, Ethical and Legal Considerations of Recaptcha "http://ieeexplore.ieee.org/document/6297942/?reload=true"