IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 22, Issue 3, Ser. IV (May - June 2020), PP 23-29
www.iosrjournals.org

CAPTCHA Bypass and Prevention Mechanisms: A Review

Mridul Wadhwa[1], Bishal K. Prasad[2], Shashank Ranjan[3],
Dr.Madhumita Kathuria[4]
Manav Rachna International Institute of Research and Studies, Haryana, India

Abstract: *Nowadays, security and privacy are considered an essential topic of concern in the digital world. While surfing the web, people are not concern about their privacy on the Internet. To retain the confidentiality of data from the world-wide-web various efforts are made, and CAPTCHA is one of them. CAPTCHAis one of the methods utilized by the application engineers to guarantee that the site administrations are not put somewhere near digital content by noxious clients. CAPTCHA depends on the Turing Test, and here the framework separates between a human interface and machine interface.*
In this paper, we analyze the various current methods of CAPTCHAs and their strength as well as weakness their vulnerabilities and different prevention, countermeasures, as well as Securing methodologies to prevent a natural CAPTCHA bypass and improve its endurance.
Keywords: *CAPTCHA BYPASS, Security, Prevention*

Date of Submission: 06-06-2020 Date of Acceptance: 22-06-2020

I. Introduction

CAPTCHA is an abbreviation for "Completely Automated Public Turing test to tell Computer and Humans Apart," which infers this is the test to tell if the client is certifiably not a malicious program. This straightforward test requires the user to retype marginally twisted images from the image.

There are a lot of dangerous individuals out there who use bots to get access. Huge organizations like Google, Facebook, PayPal are apprehensive about the security of their clients, so they have taken a few measures to ensure their information. By driving each client to breeze through this assessment, organizations are bringing down instances of losing private data into the wrong hands. Since the genesis of the Internet, individuals have attempted to mishandle sites for both games and benefits. As the maltreatment got productive, the size of misuse developed utilizing AI programming.

CAPTCHA[1] regarded as an effective way to reduce the exploitation on a website by automated bots. It oversees that the system doesn't get flooded by bots by verifying the authenticity of the user by giving a simple problem to solve, which differentiates between humans and machines covering the frontline of the underlying security of data. CAPTCHA security mechanism is critical because it can prevent malicious programs from signing up for thousands of accounts, posting many comments inweblogs.

II. Related Works

As anexponentialincrease in theuseofthe Internet byvariousgovernmentandcommercialorganizations, there had been a needtoeliminatetheillegaluseofbotsandvariousmaliciousautomatedactivities. CAPTCHA has beenthefirstlineofcontroltoeradicatesuchautomatedactivitiesforsecuringthe Internet.CAPTCHA bypassing has been achieved by using various tools like OCR, Third-party CAPTCHA Bypass Services, Segmentation, and recognition. Still, with time there is multiple CAPTCHA defeating techniques with have minimum effect on the user.

Spambot detection: An idea to design a system to automatically analyze and compare the user-submitted data and detect using filters to distinguish between human and bot by using manual identification usage of various spam filtering and detection methods. This kind of system increases the cost while CAPTCHA is comparatively cheaper.[2]

Validation: Increasing the security on the server-side by checking on the necessary fields where the various form of injection attacks occurs.

Detection: Calculation of the response time during the filling and submission as the bots are instantaneous, we can use real-world examination to differentiate between humans and bots.

Human presence to verification of data submitted by the user to provide better security and as it is almost impossible to implement so as the number of users is high, but an automatic checking addition could work for a while.[3]

Javascript code shows the presence of humans and can be used to verify by creating a tool, but it fails if the user deactivates javascript.

Deceiving bots: Usage of the Honeypot method to lure the bots into the trap by adding a web form hidden to users as bots use raw HTML. They will be unable to detect the trap, but this approach is useless in Javascript and CSS styling.[4]

As security gets complicated, so does the complexity of the attack, and it is always not about making the system immune. We can use some precautions to never let the website at a vulnerable state.

III. Captcha

CAPTCHA is a challengetestthatusesvariousmethodstoidentifywhethertheuseris humanor not, anditcomes in differenttypes. Some are easy. Some are hardertosolve, andthetestneedstobedifferenteachtime. Otherwise, someonemayusepreviousanswersforfuturetrials as well as thecodeand dataused has tobepublictobeingCAPTCHA as there are somany CAPTCHA examples. CAPTCHA can be divided into four categories:i.Text-Based, ii. Audio Based, iii.Puzzle Based, iv. Image-Based. CAPTCHA can also be sub-categorized as per the framework as follows:

- Gimpy

Gimpy is one of the prominent, trustworthy frameworks. They are made for a coordinated effort to shield the website from Bots. The gimpy is twisted text to deceive bots. The words are from word reference. Gimpy presents the test as covered material, making challenges very befuddling for Bots than real clients. Gimpy is an entirely reliable book CAPTCHA worked by CMU in a joint effort with Yahoo for its Messenger administration. Gimpy depends on the human capacity to peruse very mutilated content and the failure of the computer to do likewise. Gimpy, at that point, requests that the clients enter a subset of the words in the picture. The human client is fit for recognizing the words accurately, though a computer program can't.The disadvantage of Gimpy CAPTCHA is, Gimpy utilizations name reference words, and thus, sharp bots could be intended to check the word reference for the coordinating word.

Figure 1: GIMPY [5]

- EZ – Gimpy

A better adaptation of the Gimpy CAPTCHA developed by Yahoo. EZ – Gimpy arbitrarily picks a solitary word from a word reference and applies contortion to the content This framework is a disentangled form of gimpy. It is executed by Yahoo in their sign up. The single words are taken aimlessly from word reference and distorted to vanquish the Bots. As words are from a word reference, consequently word reference assault is defendable, but EZ-Gimpy broken by OCRs.

Figure 2: EZ-Gimpy[6]

- Baffle Text

Henry Baird created it at the University of California at Berkeley. It is a variety of Gimpy. It doesn't contain word reference words, yet it gets random letters to make hogwash yet pronounceable book. Contortionare then added to this content, and the client is asked to figure the correct word. Right now, words produced are Lucid yet don't determine any sense. Deception is added to characters in order to make it hard for

the client, and The client is then asked to take the test. This strategy facilitated the downsides in gimpy based CAPTCHA, which utilized word reference words, consequently make it progressively hard for Bots.

Figure 3: Baffle Text [7]

- reCAPTCHA

Created by Google in 2009. It was designed to work pair with their own AI. The thought behind it is fundamental. Google has begun another program to digitalize each and every examined record like books, diaries, words on photographs, and so on. At the point when AI couldn't comprehend what word is checked, it just harvests it and sends it to a client. reCAPTCHA compelled the client to offer the correct response without knowing that he helped AI to perceive content.

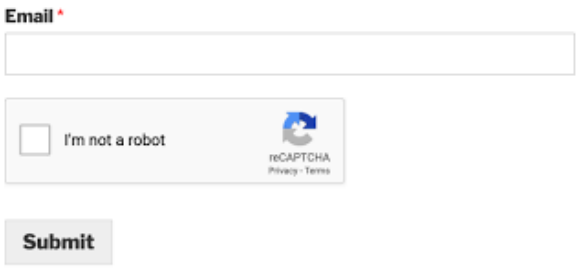

Figure 4: reCAPTCHA [8]

Image-based

Picture based CAPTCHAs are an interesting one. Here clients need to surmise similar images that are image-bearing closeness as Visual riddles. Right now, the client is given a picture, and he/she should decide the image. The comfort of these CAPTCHA frameworks is that recognition is a hard issue, and subsequently, it is hard to surpass.

Pix

It is one of the types of graphical CAPTCHA. It utilizes a database, which is of the Named picture. The images used in the pix are Pictures of material items like the female horse, work area, seat, and so on. The program is coded in a way to pick a product in arbitrary discovers four pictures aimlessly of that material from its database, distort the selected images indiscriminately, at long last put, the previous made-up to the image client for the test Asking the inquiry. Like "what are these photos of"?

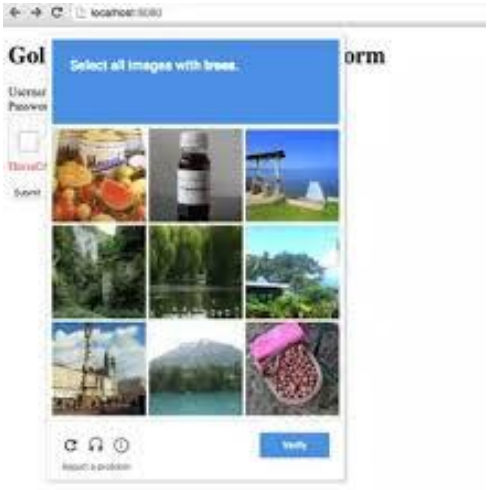

Figure 5: PIX (Image Based CAPTCHA)[9]

Puzzle based

Aside from math questions, engineers have taken a stab at utilizing straightforward questions. A case of this is an inquiry, for example, "What is the second letter of the English letters in order?" Basic queries, for the part, function admirably for Internet clients. However, concerns may, at present, emerge among specific individuals

Math solving CAPTCHA

A basic test expects you to settle some essential math exercise to go through. Even though it is a simple undertaking for everybody, you can envision how effectively it circumvents by non-human clients. Rather than a graphical code, the client needs to answer a fundamental math question. A typical model would be "2 + 2". The client at this point needs to enter the appropriate response, and on the off chance that it is right, the framework loads the following page. This alternative is useful for a great many people. Engineers ought to, regardless, guarantee that the math questions are, for the most part, fundamental ones. Math addresses that are intricate may cause troubles for people with intellectual inabilities

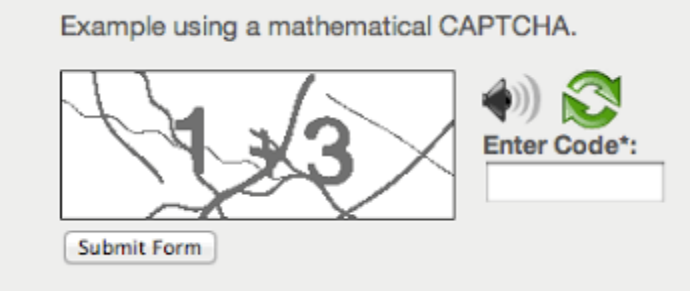

Figure 6: Maths Based [10]

Audio-based CAPTCHA

The client is given a choice to hear the Verification sound gave on screen. The client requirements to actuate a catch that would play the sound record containing the code. The client would then be able to enter the code on the given field and continue to the following procedures. This alternative is proper for daze Internet clients and those with visual disabilities. Be that as it may, hard of hearing individuals and people with cognitive disabilities may confront issues in utilizing it, the client taps on a catch to playing a sound. The page presents a Set of scores or connections containing names of sounds. The client has To tap the catch whose inscription best portrays the sound.

Figure 7: Audio BasedCAPTCHA [11]

The following were the primary type of CAPTCHA, and they can also be combined to form a better version. They also have various flaws in them discussed in the below table.

TABLE 1: Comparison of Strength, Weakness, and Vulnerabilities of different type of CAPTCHA

CAPTCHA TYPE	STRENGTH	WEAKNESS	VULNERABILITIES	COMMON METHODSOF BYPASSING
TEXT BASED	Use of distorted text image.	Misrepresentation may make it difficult for humans to read.	OCR based attack is well advanced and can achieve over 90% success of cracking.	OCR, Segmentation and recognition, Dictionary
IMAGE BASED	In the form of image-making hectic for OCR's to crack	The problem for disabled person(e.g., color-blind)Images quality may be non-recognizable	Indirect attack for obtaining the solution by using human involvement or object recognition.	Social Engineering, Random guessing, Pixelcounting,

AUDIO BASED	It is read aloud to the user for retyping.	Extra features are required for listening. Users may have a hearing impairment or visual disorder.	Social Engineering or real-time human interference	Random Guessing, Human Coercion
PUZZLE BASED	Chunked images or intellectually challenging for the client	Solving may take the time or maybe too complicated for an intellectually disabled person	Attack using a real-time human user or 3rd party problem-solving methods (e.g., CAPTCHA solving services)	Bypassing service, Human Coercion

IV. CAPTCHA Bypassing

CAPTCHA designtopreventbots, malware, andanyformofmaliciousinteractionwith a web pageas technology developed, so did the CAPTCHA types and with it evolved the bypassing methods. Some CAPTCHA Bypassing techniques are explained below.[12]

- Text-Based:

Attacks using real-time human user Relay Attack or using OCR + Dictionary are underlying vulnerabilities of text-based CAPTCHA with an accuracy of around 95-90 %. Thus, they are too primitive for the latest bypassing methods and easily avoided by bots.

- Image-Based:

We can obtain the solution from the client-side. With the use of queries in the database, it easy to guess such CAPTCHA. Attack using machine learning for object recognition, and Random guessing is very common
It is using following countermeasures like Employing encryption or code perplexity also by the use of software database creators like web crawlers. We can reduce surpassing and processing the object to be displayed in advance before subjecting them to test. But this should not match the original object and test objects.

Audio Based: Commonly bypassed by using a Social Engineering attack, which isn't accessible to countermeasure but can be stopped by better real-time detection. We can add distorted audio, making bots incapable of recognizing but enough for human.

Puzzle Based: Commonly attack using real-time human user Relay are the base problem here since most of the bots based attack do not penetrate this method. It can be done by adding communication done by user and CAPTCHA, increasing misrepresentation of letters by increasing thedistortion.

V. Prevention Techniques

CAPTCHA is at high risk from getting bypassed by multiple methods, and we need a solution not to let bots and malicious intruder to do so As various methods of CAPTCHA bypassing methods. We need a better way to secure the user data and reduce the web site to be overflowed from bots. Some Prevention techniques of CAPTCHA Bypassing are explained below.[13]

- IP Blacklisting

It is utilized to help ensure against each of the three sorts of CAPTCHA assaults, as it targets obstructing any IP address that creates suspiciously enormous volumes of traffic to a site. A site proprietor actualizes this as it would explicitly target assailants of their website. Lamentably, this shield conveys with it the hazard of blocking authentic clients who happen to share an open IP address with an assailant, are utilizing an intermediary server, or who are using an undermined machine. An IP removing administration is accessible from many webs facilitating providers8. Notwithstanding, if this alternative can't from a facilitating supplier or no facilitating supplier is included, at that point, this defense is generally mind-boggling to execute. This is because its execution would require a method for checking the source IP addresses of visits to a site just as giving server-side code to square exceptionally dynamic IP addresses (for example, by designing the Linux local firewall, IP tables).
Algorithm for checking the IP Blacklisting :
Step1: Establishment of a connection is checked.
Step2: The source address of the machine is checked if the connection gets successfully established.
Step3: Connection is checked whether it is from one of the defined local networks based on network subnet.
Step 4: Check and Match for the presence of address within the blacklist table.
Step5: It is established that the current IP, maybe malicious IP and alert, is passed.
Step6: Alert about malicious IP traffic detection is then sent to the team of network security experts where further forensics are carried out.
Step7: If found malicious, the current IP that was requesting access is banned, and blacklist tables are updated.

- Site Keys

This shield forestalls a CAPTCHA from being shown on a site other than the one proposed along these lines moderating human misuse assaults. This defense is probably not going to influence authentic clients of a place

in any capacity; all things considered, structured distinctly to forestall the showcase of a CAPTCHA on an unapproved location. When a key is given to the site proprietor, executing it might be as necessary as including a couple of lines of code.

- Response Time Monitoring

This defense ensures against bots that completely solve CAPTCHAs altogether quicker than people. As needs are, it safeguards against CAPTCHA-explaining assaults. It may be executed with the CAPTCHA supplier if observing is confined to the time required to solve the riddle. On the other hand, if the site proprietor runs the project, for example, in the structure of JavaScript downloaded to the program, it could screen the time required to finish construction and illuminate the CAPTCHA. Much the same as the past protect, a client may be contrarily influenced if utilizing an auto-fill program, as it may finish a structure quicker than a client would. One of the manners in which this protection can be executed is using customer-side scripting to compute the time slipped by from the minute information is recognized. The pace of CAPTCHA structure fruition can be determined from the contrasts between timestamps of observed occasions.

- Switching between CAPTCHAs

This shield decreases the achievement odds of bots modified to fathom explicit CAPTCHAs. It must be executed by a site proprietor, as it would ordinarily include conveying CAPTCHAs from a scope of suppliers. It may influence ease of use as the client experience .Actualizing this project is expected to be very intricate, as it would require programming a site to naturally switch between CAPTCHAs as manage various sorts of reactions relying upon the CAPTCHA supplier. It hampers the client experience very much.

- Device Fingerprinting

This protects a similar manner as IP boycotting. In any case, it has the preferred position of recognizing explicit gadgets paying little mind to regardless of whether they utilize changing IP locations, or they share IP addresses with other kind gadgets. Be that as it may, it is Or maybe progressively muddled to actualize as it requires the distinguishing proof of gadget fingerprints by gathering a scope of data about the customer stage, commonly including the OS and internet browser. Gadget fingerprinting can be actualized by the site proprietor or redistributed to a gadget fingerprinting administration supplierReal clients may be influenced on the off chance that they were erroneously recognized as bots and accordingly obstructed from collaborating with the site or tested with more CAPTCHA puzzles.
client is a framework written in the Javascript fingerprints device visiting the website, the checks are performed, and output is Boolean.
Algorithm to Check End Side user is Human or Bot:
Step1: Check browser type(Mobile/Desktop Application)
Step2: IF True: return Browser type, Browser version, Browser Major Version, and Browser Name.
Step3: Check Operating System
Step4: IF True: OS name, OS version.
Step5: check CPU
Step6: IF True, return CPU info.
Step7: Check Client is Mobile
Step8: IF True: return DeviceName, OS version, BrowserName.
Step9: Return screenResolution, colorDepth, currentResolution, availableResolution, DeviceXDPI, Device YDPI.
Step10: Check Plugins
Step11: IF TRUE, return PluginType, PluginVersion.
Step12: Check Region and Timezone
Step13: IF TRUE return Region, Timezone, Local Location, System Language.

- Brand Customization

This shield will caution clueless clients that they are being tricked into unraveling a CAPTCHA puzzle of another site. This defense must be actualized if the CAPTCHA supplier gives a customization choice to the site proprietor. The expansion of a web logo, and maybe a little measure of caution content, is probably not going to have a critically antagonistic impact on clients. Since this is reliant on the CAPTCHA supplier, actualizing it would not usually require a site proprietor to do any more than finding a way to show the modified CAPTCHA.

TABLE 2: Comparison of various security methods

METHODS	RESTRICTS	IMPLEMENTATION	AFFECT ON USER
IP BLACKLISTING	ALL ATTACKS	COMPLEX	HIGH
SITE KEYS	HUMAN EXPLOITATION	SIMPLE	LOW
RESPONSE MONITORING TIME	CAPTCHA SOLVING	MODERATE	LOW
SWITCHING BETWEEN CAPTCHA	CAPTCHA SOLVING	COMPLEX	HIGH
DEVICE FINGERPRINTING	ALL ATTACKS	COMPLEX	LOW
BRAND CUSTOMIZATION	HUMAN EXPLOITATION	SIMPLE	LOW

VI. Conclusion

We conclude by quoting the platitude Prevention is Better than Cure remains constant likewise in the field of security and privacy. It's essential to improve Internet cleanliness to reduce vulnerabilities in web browsers. CAPTCHAs are powerless against assaults. A supposed decent CAPTCHA plan can break with a by and large (division and afterward acknowledgment) achievement pace of over 60%. In this way, we find that CAPTCHAs give less security. Regardless of whether division opposition is a sound guideline for structuring secure content based CAPTCHAs, it is essential to ensure that a plan can't go to any known (and preferably obscure) division strategy. Planning CAPTCHAs that show both high power and convenience is a lot harder than it may give off an impression of being because present aggregate comprehension of this theme is little and the necessities, apparatuses, and strategies for surveying CAPTCHA structures are nearly nil. Numerous sites have utilized CAPTCHAs to sift through associations by bots. Be that as it may, assailants. Have discovered approaches to dodge CAPTCHAs by programming bots to comprehend or sidestep them, or even hand-off them for people to settle. To diminish the odds of achievement of such assaults, The expansion of specific shields can fortify CAPTCHAs. Right now, talk about seven existing protections just as five novel shields intended to make evading. These shields are not fundamentally unrelated and can add numerous layers of insurance to a CAPTCHA. We further give an elevated level examination of their adequacy in We are tending to the risk presented by CAPTCHA-overcoming procedures. To concentrate on shields that are usable, we confine our thoughtfulness regarding those which have an insignificant unfavorable impact on the client experience as well as protecting their data.

References

[1]. A. L. Coates, H. S. Baird, and R. J. Fateman, "PessimalPrint: a reverse Turing test," *International Journal on Document Analysis and Recognition*, vol. 5, pp. 158-163, 2003.

[2]. J. Yan, "Bot, cyborg and automated turing test," in *Security Protocols Workshop*, 2006, pp. 190-197.

[3]. Buckler C. 10 things to check before using a CAPTCHA. Available at: http://www.sitepoint.com/ captcha-alternatives/ [accessed December 2012]

[4]. Keeping out the bad bots. Available at: http://www.timezoneoneblog.com/2013/03/04/keeping-out-thebad-bots/ [accessed December 2012]

[5]. C. Pope and K. Kaur, "Is it human or computer? Defending e-commerce with CAPTCHAs," *IT professional*, vol. 7, pp. 43-49, 2005.

[6]. M. Blum, L. Von Ahn, J. Langford, and N. Hopper, "The CAPTCHA project,"Completely automatic public turing test to tell computers and humans apart,"" *Dept. of Computer Science, Carnegie-Mellon University, www. captcha. net*, 2000.

[7]. Baffletext: A Human Interactive Proof. . Proceedings of SPIE (pp. 305-316). SPIE . Coates, A. L., Henry, S. B., &Fateman, R. J. (2001).

[8]. Von. Ahn L. 2009. Human Computation (reCAPTCHA). 46thACM/ IEE Design Automation

[9]. http://www.captcha.net/captchas/pix (Accessed on 10December 2015).Conference. pp. 418-419.

[10]. H. Gao, D. Yao, H. Liu, X. Liu, and L. Wang, "A Novel Image Based CAPTCHA Using Jigsaw Puzzle," in *13th IEEE International Conference on Computational Science and Engineering (CSE)*, 2010, pp. 351-356.

[11]. Computer Vision and Pattern Recognition (CVPR '03) (pp. 134– 144). IEEE . Moy, G., Jones, N., Harkless, C., & Potter, R. (2004). Distortion Estimation Techniques in Solving Visual CAPTCHAs

[12]. CAPTCHA: Attacks and Weaknesses against OCR Technology. Global Journal of Computer Science and Technology Neural & Artificial Intelligence,13. Chellapilla, K., & Simard, P. Y. (2004).

[13]. BYPASSING CAPTCHA BY MACHINE—A PROOF FOR PASSING THE TURING TEST Ahmad B. A. Hassanat, PhD IT Department, Mu'tah University, Jordan (2014).

Mridul Wadhwa, et. al. "CAPTCHA Bypass and Prevention Mechanisms: A Review." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22(3), 2020, pp. 23-29