A detailed study of CAPTCHA

Abstract:

You will be browsing Google before suddenly being asked to prove that you are not a robot. As a result, you will need to enter the correct characters as shown to you on screen or identify which pictures contain traffic lights or something similar. Only then will you be allowed to carry on with your search. This is based on captcha technology. The main purpose of this article was to explore the application of captcha on a website and to get the information on its relevance. The interpretations and conclusion of this article briefly explain its history, modified form, its types, working style, alternatives.

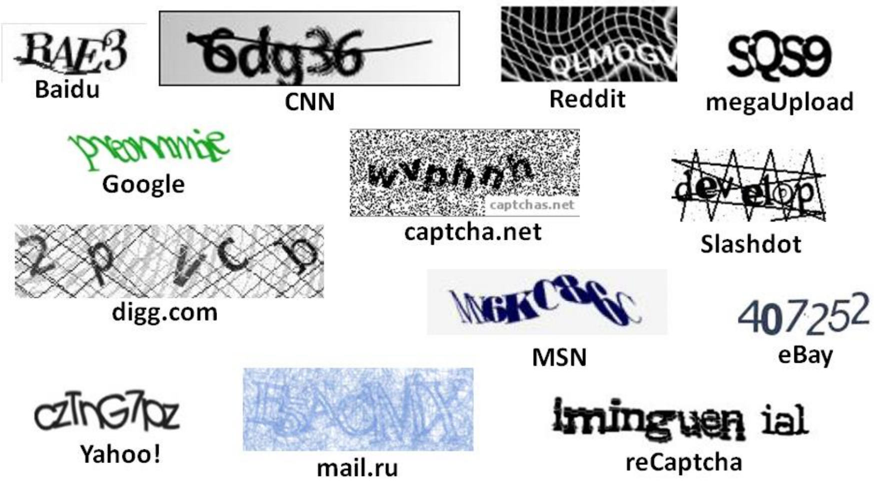
1. Introduction:

Captcha was developed by engineers at Carnegie Mellon University in early 2000. The team was led by Luis von Ahn. The term “CAPTCHA” also known as (Completely Automated Public Turing test to tell Computers and Human Apart) was introduced due to several reasons. One of the reasons was the failure to keep up the large database of various websites. Similarly, millions of accounts were being created in Yahoo every second through programming. To control that, it was necessary to identify whether the person creating the account was a human or a script (programming code). CAPTCHA became the way to get rid of all these problems. It featured different letters and numbers. Which is rotated from different angles. People can easily recognize such letters if they look at them carefully, but no computer or any programming code can recognize the same thing. From that, it could be determined that the person filling in the form or opening the account on the website is the same person. Similarly, the reCAPTCHA was designed to give the captcha a more sophisticated look. With the intention of showing something more meaningful than the unnecessary things shown in the captcha, the captcha developer started scanning the various books available. Google bought ReCAPTCHA in 2009 and took ownership of it after similar problems began to appear from Google search engine to Gmail. Since then, Google has produced various ideas on how to make ReCAPTCHA better than ever before.

2. Types of CAPTCHA methods

2.1. Text-Based CAPTCHAs

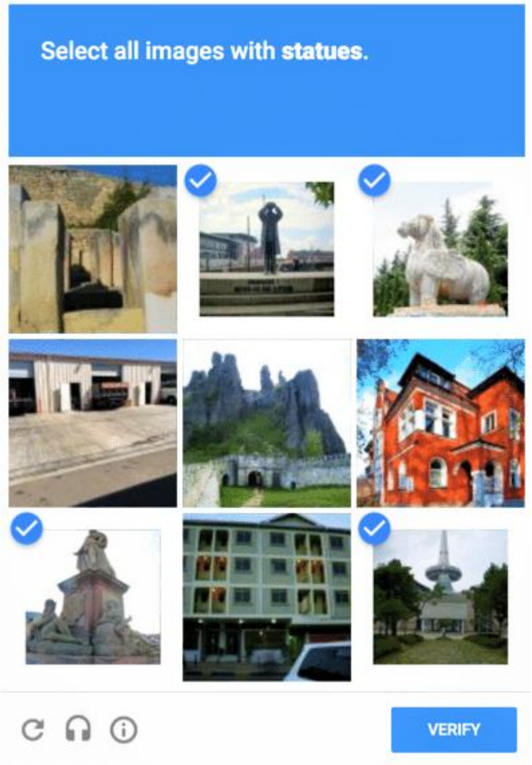
The most standard kind of check is text CAPTCHAs. These CAPTCHAs can fuse prominent articulations or explanations, similarly as uncommon digits, and letter blends. Some substance-based CAPTCHAs break down distinct kinds of capitalization. These characters are shown in an odd style by the CAPTCHA, requiring translation. Strategies for making text-based CAPTCHAs include Gimpy, EZ-Gimpy, Gimpy-r, Simard’s HIP.



Page Break

2.2. Image-Based CAPTCHAs

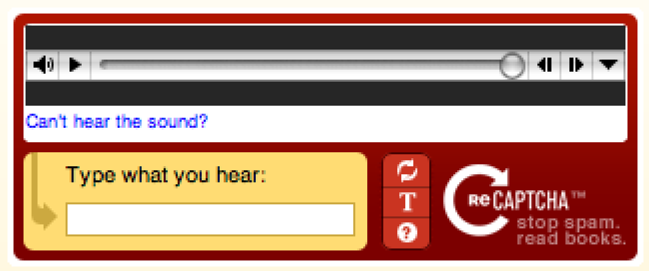
Manual human tests utilizing pictures are developed in a split second clear graphical components instead of a vexing strategy including digits and letters. Finally, a few photographs of ordinary things are compared. The customer should feature which photographs have all the earmarks of being the most significant or show which ones tackle a semantic issue. Google, then again, utilizes Google Street View CAPTCHAs that expect clients to enter a street address or a road sign into the material box. Most clients can address an image-based CAPTCHA very quickly. Regardless, a PC program's capacity to acquire an addressed picture, then, at that point request it is anything but, and afterward work out near one is restricted partly. Thus, picture-based CAPTCHAs give preferred security over text-based cycles.



Select all images with **statues.**

2.3. Audio-Based CAPTCHAs

Manual human tests are a development that permits individuals to get to obstructed sites. These CAPTCHAs are as often as possible utilized related to message based and picture-based CAPTCHAs. Customers ought to expect a progression of moving characters or numbers in a decent CAPTCHA. Bots cannot separate crucial characters from establishment shock in these CAPTCHAs. Concerning bots, these mechanical gatherings, like substance-based CAPTCHAs, can be difficult for individuals to fathom.

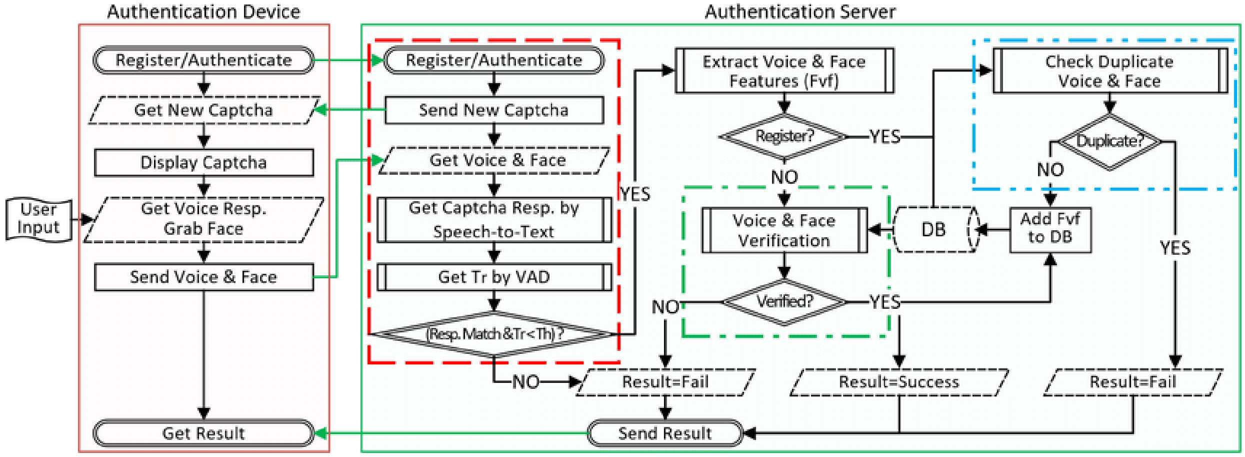


Page Break

3.Working style

A Web server may be holding both public and protected resources that may be in the form of web pages, data stored in a database or files, or some other service intended to be used by human users on the client. User requests for a resource are sent by the client computer to the server, which is granted to it if the resource is not protected. In case the resource is CAPTCHA protected, access is granted to it only after passing the CAPTCHA test as depicted in the figure. The server uses some CAPTCHA image generation algorithms to generate a CAPTCHA image.

Different CAPTCHA techniques use different algorithms for image generation which may employ the use of images stored in an image database. The state information along with the Global Unique Identifier (GUID) of the client and the CAPTCHA solution is stored in the State Information Database (SID) on the server. Storing the GUID of the client ensures that only the client that received CAPTCHA can produce a valid solution. Instead of storing the CAPTCHA solution and other state information on the server in SID, it may be stored in hashed or encrypted form in a cookie on the client. A web page containing the generated CAPTCHA image and the cookie is posted to the client which renders it in a web browser to the user. A human operator responds to the CAPTCHA test and the response is passed by the client to the server. The server verifies the authenticity of the CAPTCHA solution by comparing the stored GUID and the GUID of the client sending the solution. The solution provided by the client is next compared with the solution stored in SID or cookie and accordingly, either access is granted or denied. In case access is denied, a message is posted to the client and the process starts afresh. A CAPTCHA implementation may temporarily block access for a client if it repeatedly fails to respond to a number of CAPTCHA tests. Further, for a particular session once a CAPTCHA challenge has been passed by a client, subsequent access to protected resources on the server may be granted to it without putting it to further tests.



4.Circumvention

Although CAPTCHA is a valuable tool for protecting against bots and automated hacking tools, there is still a flaw in the system through which it can easily get bypassed. Howard Yeend has identified two implementation issues with poorly designed CAPTCHA systems:[1]

- Some CAPTCHA protection systems can be bypassed without using OCR simply by reusing the session ID of a known CAPTCHA image
- CAPTCHAs residing on shared servers also present a problem; a security issue on another virtual host may leave the CAPTCHA issuer's site vulnerable.

[4.1.] **There have been some notable attacks reported against various CAPTCHA schemes:**

- Mori et al. published a paper in IEEE CVPR'03 detailing a method for defeating one of the most popular CAPTCHAs, EZ-Gimpy, which was tested as being 92% accurate in defeating it. The same method was also shown to defeat the more complex and less-widely deployed Gimpy program 33% of the time. However, the existence of implementations of their algorithm in actual use is indeterminate currently.
- PWNtcha has made considerable progress in defeating commonly used CAPTCHAs, which has contributed to a general migration towards more sophisticated CAPTCHAs.
- Podec, a trojan discovered by the security company Kaspersky, forwards CAPTCHA requests to an online human translation service that converts the image to text, fooling the system. Podec targets Android mobile devices.

5. Alternative CAPTCHA Schemes

With the demonstration that text distortion-based CAPTCHAs are vulnerable to machine learning-based attacks, some researchers have proposed alternatives including image recognition CAPTCHAs which require users to identify simple objects in the images presented. The argument in favor of these schemes is that tasks like object recognition are typically more complex to perform than text recognition and therefore should be more resilient to machine learning-based attacks. Here are some of the notable alternative CAPTCHA schemes:

- Chew et al. published their work in the 7th International Information Security Conference, ISC'04, proposing three different versions of image recognition CAPTCHAs, and validating the proposal with user studies. It is suggested that one of the versions, the anomaly CAPTCHA, is best with 100% of human users being able to pass an anomaly CAPTCHA with at least 90% probability in 42 seconds.
- Datta et al. published their paper at the ACM Multimedia'05 Conference, named IMAGINATION (Image Generation for Internet Authentication), proposing a systematic way to image recognition CAPTCHAs. Images are distorted in such a way that state-of-the-art image recognition approaches (which are potential attack technologies) fail to recognize them.
- Microsoft (Jeremy Elson, John R. Douceur, Jon Howell, and Jared Saul) claim to have developed Animal Species Image Recognition for Restricting Access (ASIRRA) which asks users to distinguish cats from dogs. Microsoft had a beta version of this for websites to use. They claim "Asirra is easy for users; it can be solved by humans 99.6% of the time in under 30 seconds. Anecdotally, users seemed to find the experience of using Asirra much more enjoyable than a text-based CAPTCHA." This solution was described in a 2007 paper in the Proceedings of the 14th ACM Conference on Computer and Communications Security

(CCS). However, this project was closed in October 2014 and is no longer available.

References

1. https://en.wikipedia.org/wiki/CAPTCHA.Retrieved on 2022-06-07.

2. https://www.g2.com/articles/captcha. Retrieved on 2022-06-05.

3. https://www.wallarm.com/what/what-is-captcha-types-and-examples. Retrieved in 2022-06-06.

4. https://www.cloudflare.com/learning/bots/how-captchas-work/. Retrieved in 2022-06-06.

5. “rtCaptcha: A Real-Time Captcha Based Liveness Detection System” (PDF). Erkam Uzun, Simon Pak Ho Chung, Irfan Essa and Wenke Lee. Published in February 2018.

6. M.Tariq Bandey and N.A. Shah. “A Study of CAPTCHAs for Securing Web Services” (PDF). Archived from the original pdf in December 2011.

7. https://www.researchgate.net. Retrieved on 2022-06-07.